PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/54, 15/55, 9/12, 9/16, C07K 14/705, C12N 15/11, C07K 16/40, 16/28, C12N 5/12, 15/62, C12Q 1/42, 1/48

A2

(11) International Publication Number: WO 98/4931

(43) International Publication Date: 5 November 1998 (05.11.9

(21) International Application Number: PCT/US98/08439

(22) International Filing Date: 27 April 1998 (27.04.98)

(30) Priority Data:

60/044,428	28 April 1997 (28.04.97)	US
60/047,222	20 May 1997 (20.05.97)	US
60/049,756	11 June 1997 (11.06.97)	US
60/049,477	11 June 1997 (11.06.97)	US
60/049,914	18 June 1997 (18.06.97)	US
60/063,595	23 October 1997 (23.10.97)	US

(71) Applicant (*for all designated States except US*): SUGEN, INC. [US/US]; 351 Galveston Drive, Redwood City, CA 94063 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): PLOWMAN, Greg, D. [US/US]; 4 Honeysuckle Lane, San Carlos, CA 94070 (US). CLARY, Douglas [US/US]; 164 Midcrest Way, San Francisco, CA 94131 (US). JALLAL, Bahija [MA/US]; 101 O'Keefe Street, Melo Park, CA 94025 (US). PELES, Elior [IL/IL]; Hanasi Harishon 51, 76303 Rhovot (IL). ONRUST, Susan [US/NZ]; 6 Summit Drive, Mt. Albert, Auckland 3 (NZ). MARKBY, Dave [US/US]; Apartment A, 477 Burnett Avenue, San Francisco, CA 94131 (US). COURTNEIDC Sara, A. [GB/US]; 1408 Alvarado Avenue, Burlingame, C 94010 (US). APP, Harald [DE/US]; 630 27th Street, S Francisco, CA 94131 (US). HUI, Terance, H. [CN/US]; 3 Skyline Drive, Daly City, CA 94015 (US).

(74) Agents: WARBURG, Richard, J. et al.; Lyon & Lyon LL Suite 4700, 633 West Fifth Street, Los Angeles, 90071-2066 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, I BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, M MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, AR patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eura patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), Europ patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republi upon receipt of that report.

(54) Title: DIAGNOSIS AND TREATMENT OF PHOSPHATASE OR KINASE-RELATED DISORDERS

(57) Abstract

The present invention relates to phosphatases and kinases, nucleic acids encoding such polypeptides, cells, tissues and an containing such nucleic acids, antibodies to such polypeptides, assays utilizing such polypeptides, and methods relating to all of the foreg Methods for treatment, diagnosis, and screening are provided for phosphatase or kinase related diseases or conditions characterized b abnormal interaction between a phosphatase or a kinase and its binding partner.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DESCRIPTION

Diagnosis And Treatment Of Tyrosine Phosphatase-Related Disorders And Related Methods

Field Of The Invention

The present invention relates to protein tyrosine phosphatases. In particular, the invention concerns proteins we have named PTP04, SAD, PTP05, PTP10, ALP, and ALK-7, nucleotide sequences encoding these proteins, and various products and assay methods that can be used for identifying compounds useful for the diagnosis and treatment of various diseases and conditions related to these proteins, for example cell proliferative disorders.

Background Of The Invention

The following description is provided to aid in understanding the invention but is not admitted to be prior art to the invention.

Cellular signal transduction is a fundamental mechanism whereby external stimuli that regulate diverse cellular processes are relayed to the interior of cells. One of the key biochemical mechanisms of signal transduction involves the reversible phosphorylation of proteins, which enables regulation of the activity of mature proteins by altering their structure and function. The best characterized protein kinases in eukaryotes phosphorylate proteins on the alcohol moiety of serine, threonine and tyrosine residues. These kinases largely fall into two groups, those specific for phosphorylating serines and threonines, and those specific for phosphorylating tyrosines.

The phosphorylation state of a given substrate is also regulated by a class of proteins responsible for removal of the

phosphate group added to a given substrate by a protein kinase. The protein phosphatases can also be classified as being specific for either serine/threonine or tyrosine. The known enzymes can be divided into two groups - receptor and non-receptor type proteins. Most receptor-type protein tyrosine phosphatases (RPTPs) contain two conserved catalytic tyrosine phosphatase domains each of which encompasses a segment of 240 amino acid residues (Saito et al, Cell Growth and Diff. 2:59-65, 1991). The RPTPs can be subclassified further based upon the amino acid sequence diversity of their extracellular domains (Saito, et al, supra; Krueger, et al, Proc. Natl. Acad. Sci. USA 89:7417-7421, 1992). Alignment of primary peptide sequences of both types of known PTPases shows some sequence consensus in catalytic domains and has made it possible to identify cDNAs encoding proteins with tyrosine phosphate activity via the polymerase chain reaction (PCR).

Many kinases and phosphatases are involved in regulatory cascades wherein their substrates may include other kinases and phosphatases whose activities are regulated by their phosphorylation state. Ultimately the activity of some downstream effector is modulated by phosphorylation resulting from activation of such a pathway.

It is well established that the abnormal or inappropriate activity of tyrosine kinases and/or tyrosine phosphatases plays a role in a variety of human disorders including cell proliferative disorders such as cancer, fibrotic disorders, disorders of the immune system and metabolic disorders such as diabetes. A need, therefore, exists to identify new tyrosine kinases and phosphatases as a first step in understanding a disease process and the subsequent identification of therapeutic treatments for the disorder.

Summary Of The Invention

The present invention concerns PTP04, SAD, PTP05, PTP10, Alp, and ALK-7 polypeptides, nucleic acids encoding such polypeptides, cells, tissues and animals containing such nucleic acids, antibodies to the polypeptides, assays utilizing the polypeptides, and methods relating to all of the foregoing.

A first aspect of the invention features an isolated, enriched, or purified nucleic acid molecule encoding a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide.

By "isolated" in reference to nucleic acid is meant a polymer of 14, 17, 21 or more nucleotides conjugated to each other, including DNA or RNA that is isolated from a natural source or that is synthesized. The isolated nucleic acid of the present invention is unique in the sense that it is not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular (*i.e.*, chromosomal) environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only nucleotide sequence present, but that it is essentially free (about 90 - 95% pure at least) of non-nucleotide material naturally associated with it and thus is meant to be distinguished from isolated chromosomes.

By the use of the term "enriched" in reference to nucleic acid is meant that the specific DNA or RNA sequence constitutes a significantly higher fraction (2 - 5 fold) of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other DNA or RNA present, or by a preferential increase in the amount of the specific DNA or RNA sequence, or by a combination of the two.

However, it should be noted that "enriched" does not imply that there are no other DNA or RNA sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term "significant" here is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other nucleic acids of about at least 2 fold, more preferably at least 5 to 10 fold or even more. The term also does not imply that there is no DNA or RNA from other sources. The other source DNA may, for example, comprise DNA from a yeast or bacterial genome, or a cloning vector such as pUC19. This term distinguishes the sequence from naturally occurring enrichment events, such as viral infection, or tumor type growths, in which the level of one mRNA may be naturally increased relative to other species of mRNA. That is, the term is meant to cover only those situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that a nucleotide sequence be in purified form. The term "purified" in reference to nucleic acid does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment (compared to the natural level this level should be at least 2-5 fold greater, *e.g.*, in terms of mg/mL). Individual clones isolated from a cDNA library may be purified to electrophoretic homogeneity. The claimed DNA molecules obtained from these clones can be obtained directly from total DNA or from total RNA. The cDNA clones are not naturally occurring, but rather are preferably obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The construction of a cDNA library from mRNA involves the creation of a synthetic substance (cDNA) and pure

individual cDNA clones can be isolated from the synthetic library by clonal selection of the cells carrying the cDNA library. Thus, the process which includes the construction of a cDNA library from mRNA and isolation of distinct cDNA clones yields an approximately 10^6-fold purification of the native message. Thus, purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. The term is also chosen to distinguish clones already in existence which may encode PTP04, SAD, PTP05, PTP10, ALP, or Alk-7 but which have not been isolated from other clones in a library of clones. Thus, the term covers clones encoding PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 which are isolated from other non-PTP04, non-SAD, non-PTP05, non-PTP10, non-ALP, or non-ALK-7 clones.

A PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide can be encoded by a full-length nucleic acid sequence or any portion of the full-length nucleic acid sequence. In preferred embodiments the isolated nucleic acid comprises, consists essentially of, or consists of a nucleic acid sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8, a nucleic acid sequence that hybridizes to the nucleic acid sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8 or a functional derivative (as defined below) of either. The nucleic acid may be isolated from a natural source by cDNA cloning or subtractive hybridization; the natural source may be mammalian (human) blood, semen, or tissue and the nucleic acid may be synthesized by the triester or other method or by using an automated DNA synthesizer.

The term "hybridize" refers to a method of interacting a nucleic acid sequence with a DNA or RNA molecule in solution or

on a solid support, such as cellulose or nitrocellulose. If a nucleic acid sequence binds to the DNA or RNA molecule with high affinity, it is said to "hybridize" to the DNA or RNA molecule. The strength of the interaction between the probing sequence and its target can be assessed by varying the stringency of the hybridization conditions. Various low or high stringency hybridization conditions may be used depending upon the specificity and selectivity desired. Stringency is controlled by varying salt or denaturant concentrations.

As a general guideline, high stringency conditions (hybridization at 50-65 °C, 5X SSPC, 50% formamide, wash at 50-65 °C, 0.5X SSPC) can be used to obtain hybridization between nucleic acid sequences having regions which are greater than about 90% complementary. Low stringency conditions (hybridization at 35-37 °C, 5X SSPC, 40-45% formamide, wash at 42 °C SSPC) can be used so that sequences having regions which are greater than 35-45% complementarity will hybridize to the probe. These conditions only represent examples of stringency conditions and those skilled in the art recognize that these conditions may be changed depending on the particular mode of practice. Further examples of hybridization conditions are shown in the examples below. Those skilled in the art will recognize how such conditions can be varied to vary specificity and selectivity. Under highly stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having one or two mismatches out of 20 contiguous nucleotides.

In yet other preferred embodiments the nucleic acid is an isolated conserved or unique region, for example those useful for the design of hybridization probes to facilitate identification and cloning of additional polypeptides, or for the

design of PCR probes to facilitate cloning of additional polypeptides.

By "conserved nucleic acid regions", it is meant regions present on two or more nucleic acids encoding a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide, to which a particular nucleic acid sequence can hybridize under lower stringency conditions. Examples of lower stringency conditions suitable for screening for nucleic acids encoding PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptides are provided in Abe, et al. J. Biol. Chem. 19:13361 (1992). Preferably, conserved regions differ by no more than 5 out of 20 continguous nucleotides.

By "unique nucleic acid region" it is meant a sequence present in a full length nucleic acid coding for a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide that is not present in a sequence coding for any other known naturally occurring polypeptide. Such regions preferably comprise 14, 17, 21 or more contiguous nucleotides present in the full length nucleic acid encoding a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide. In particular, a unique nucleic acid region is preferably of human origin.

The invention also features a nucleic acid probe for the detection of a nucleic acid encoding a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide in a sample. The nucleic acid probe contains nucleic acid that will hybridize specifically to a sequence of at least 14, preferably 17, 20 or 22, continguous nucleotides set forth in SEQ ID NO:1 or a functional derivative thereof. The probe is preferably at least 14, 17 or more bases in length and selected to hybridize specifically to a unique region of a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 endocing nucleic acid.

In preferred embodiments the nucleic acid probe hybridizes to nucleic acid encoding at least 14 contiguous amino acids of

the full-length sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8 or a functional derivative thereof. Various low or high stringency hybridization conditions may be used depending upon the specificity and selectivity desired. Under highly stringent hybridization conditions only highly complementary nucleic acid sequences hybridize. Preferably, such conditions prevent hybridization of nucleic acids having 1 or 2 mismatches out of 20 contiguous nucleotides.

Methods for using the probes include detecting the presence or amount of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 RNA in a sample by contacting the sample with a nucleic acid probe under conditions such that hybridization occurs and detecting the presence or amount of the probe bound to PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 RNA. The nucleic acid duplex formed between the probe and a nucleic acid sequence coding for a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide may be used in the identification of the sequence of the nucleic acid detected (for example see, Nelson et al., in Nonisotopic DNA Probe Techniques, p. 275 Academic Press, San Diego (Kricka, ed., 1992)). Kits for performing such methods may be constructed to include a container means having disposed therein a nucleic acid probe.

The invention also features recombinant nucleic acid, preferably in a cell or an organism. The recombinant nucleic acid may contain a sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8 or a functional derivative thereof and a vector or a promoter effective to initiate transcription in a host cell. The recombinant nucleic acid can alternatively contain a transcriptional initiation region functional in a cell, a sequence complimentary to an RNA sequence encoding a

PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide and a transcriptional termination region functional in a cell.

Another aspect of the invention features an isolated, enriched, or purified PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide.

By "PTP04 polypeptide" it is meant an amino acid sequence substantially similar to the sequence shown in SEQ ID NO:9, or fragments thereof. By "SAD polypeptide" it is meant an amino acid sequence substantially similar to the sequence shown in SEQ ID NO:10, or fragments thereof. By "PTP05 polypeptide" or "PTP10 polypeptide" it is meant an amino acid sequence substantially similar to the sequence shown in SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or SEQ ID NO:14, or fragments thereof. By "ALP polypeptide" it is meant an amino acid sequence substantially similar to the sequence shown in SEQ ID NO:15, or fragments thereof. By "ALK-7 polypeptide" it is meant an amino acid sequence substantially similar to the sequence shown in SEQ ID NO:16, or fragments thereof. Two substantially similar sequences will preferably have at least 90% identity (more preferably at least 95% and most preferably 99-100%) to each other.

By "identity" is meant a property of sequences that measures their similarity or relationship. Identity is measured by dividing the number of identical residues in the two sequences by the total number of residues and multiplying the product by 100. Thus, two copies of exactly the same sequence have 100% identity, but sequences that are less highly conserved and have deletions, additions, or replacements have a lower degree of identity. Those skilled in the art will recognize that several computer programs are available for determining sequence identity.

By "isolated" in reference to a polypeptide is meant a polymer of 6, 12, 18 or more amino acids conjugated to each

other, including polypeptides that are isolated from a natural source or that are synthesized. The isolated polypeptides of the present invention are unique in the sense that they are not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only amino acid chain present, but that it is essentially free (about 90 - 95% pure at least) of material naturally associated with it.

By the use of the term "enriched" in reference to a polypeptide it is meant that the specific amino acid sequence constitutes a significantly higher fraction (2 - 5 fold) of the total of amino acids present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by a person by preferential reduction in the amount of other amino acids present, or by a preferential increase in the amount of the specific amino acid sequence of interest, or by a combination of the two. However, it should be noted that "enriched" does not imply that there are no other amino acid sequences present, just that the relative amount of the sequence of interest has been significantly increased. The term significant here is used to indicate that the level of increase is useful to the person making such an increase, and generally means an increase relative to other amino acids of about at least 2 fold, more preferably at least 5 to 10 fold or even more. The term also does not imply that there is no amino acid from other sources. The other source amino acid may, for example, comprise amino acid encoded by a yeast or bacterial genome, or a cloning vector such as pUC19. The term is meant to cover only those

situations in which a person has intervened to elevate the proportion of the desired nucleic acid.

It is also advantageous for some purposes that an amino acid sequence be in purified form. The term "purified" in reference to a polypeptide does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the sequence is relatively purer than in the natural environment (compared to the natural level this level should be at least 2-5 fold greater, *e.g.*, in terms of mg/mL). Purification of at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. The substance is preferably free of contamination at a functionally significant level, for example 90%, 95%, or 99% pure.

In another aspect the invention features an isolated, enriched, or purified PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide fragment.

By "a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide fragment" it is meant an amino acid sequence that is less than the full-length PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 amino acid sequence shown in SEQ ID NO:2. Examples of fragments include PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 domains, PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 mutants and PTP04-, SAD-, PTP05-, PTP10-, ALP-, or ALK-7-specific epitopes.

By "a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 domain" it is meant a portion of the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide having homology to amino acid sequences from one or more known proteins wherein the sequence predicts some common function, interaction or activity. Well known examples of domains are the SH2 (Src Homology 2) domain (Sadowski, et al, Mol. Cell. Biol. 6:4396, 1986; Pawson and Schlessinger, Curr. Biol. 3:434, 1993), the SH3 domain (Mayer, et al, Nature 332:272, 1988; Pawson and Schlessinger, Curr.

Biol. 3:434, 1993), and pleckstrin (PH) domain (Ponting, TIBS 21:245, 1996; Haslam, et al, Nature 363:309, 1993), all of which are domains that mediate protein:protein interaction, and the kinase catalytic domain (Hanks and Hunter, FASEB J 9:576-595, 1995). Computer programs designed to detect such homologies are well known in the art. The relative homology is at least 20%, more preferably at least 30% and most preferably at least 35%.

By "a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 mutant" it is meant a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide which differs from the native sequence in that one or more amino acids have been changed, added or deleted. Changes in amino acids may be conservative or non-conservative. By "conservative" it is meant the substitution of an amino acid for one with similar properties such as charge, hydrophobicity, structure, etc. Examples of polypeptides encompassed by this term include, but are not limited to, (1) chimeric proteins which comprise a portion of a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide sequence fused to a non-PTP04, a non-SAD, a non-PTP05, a non-PTP10, a non-ALP, or a non-ALK-7 polypeptide sequence, for example a polypeptide sequence of hemagglutinin (HA), (2) PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 proteins lacking a specific domain, for example the catalytic domain, and (3) PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 proteins having a point mutation. A PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 mutant will retain some useful function such as, for example, binding to a natural binding partner, catalytic activity, or the ability to bind to a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 specific antibody (as defined below).

By "PTP04-, SAD-, PTP05-, PTP10-, ALP-, or ALK-7-specific epitope" it is meant a sequence of amino acids that is both antigenic and unique to PTP04, SAD, PTP05, PTP10, ALP, or ALK-

7. PTP04-, SAD-, PTP05-, PTP10-, ALP-, or ALK-7-specific epitope can be used to produce PTP04-, SAD-, PTP05-, PTP10-, ALP-, or ALK-7-specific antibodies, as more fully described below. Particularly preferred epitopes are shown in Examples below.

By "recombinant PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide" it is meant to include a polypeptide produced by recombinant DNA techniques such that it is distinct from a naturally occurring polypeptide either in its location (*e.g.*, present in a different cell or tissue than found in nature), purity or structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

In yet another aspect the invention features an antibody (*e.g.*, a monoclonal or polyclonal antibody) having specific binding affinity to a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide or polypeptide fragment. By "specific binding affinity" is meant that the antibody binds to target polypeptide with greater affinity than it binds to other polypeptides under specified conditions. Antibodies or antibody fragments are polypeptides which contain regions that can bind other polypeptides. The term "specific binding affinity" describes an antibody that binds to a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide with greater affinity than it binds to other polypeptides under specified conditions.

The term "polyclonal" refers to antibodies that are heterogenous populations of antibody molecules derived from the sera of animals immunized with an antigen or an antigenic functional derivative thereof. For the production of polyclonal antibodies, various host animals may be immunized by injection with the antigen. Various adjuvants may be used to increase the immunological response, depending on the host species.

"Monoclonal antibodies" are substantially homogenous populations of antibodies to a particular antigen. They may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. Monoclonal antibodies may be obtained by methods known to those skilled in the art. See, for example, Kohler, et al., *Nature* 256:495-497 (1975), and U.S. Patent No. 4,376,110.

The term "antibody fragment" refers to a portion of an antibody, often the hypervariable region and portions of the surrounding heavy and light chains, that displays specific binding affinity for a particular molecule. A hypervariable region is a portion of an antibody that physically binds to the polypeptide target.

Antibodies or antibody fragments having specific binding affinity to a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide may be used in methods for detecting the presence and/or amount of a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide in a sample by probing the sample with the antibody under conditions suitable for formation of an immunocomplex between the antibody and the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide and detecting the presence and/or amount of the antibody conjugated to the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide. Diagnostic kits for performing such methods may be constructed to include antibodies or antibody fragments specific for PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 as well as a conjugate of a binding partner of the antibodies or the antibodies themselves.

An antibody or antibody fragment with specific binding affinity to a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide can be isolated, enriched, or purified from a prokaryotic or eukaryotic organism. Routine methods known to those skilled in the art enable production of antibodies or antibody fragments, in both prokaryotic and eukaryotic

organisms. Purification, enrichment, and isolation of antibodies, which are polypeptide molecules, are described above.

In another aspect the invention features a hybridoma which produces an antibody having specific binding affinity to a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide. By "hybridoma" is meant an immortalized cell line which is capable of secreting an antibody, for example a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 antibody. In preferred embodiments the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 antibody comprises a sequence of amino acids that is able to specifically bind a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide.

In another embodiment, the invention encompasses a recombinant cell or tissue containing a purified nucleic acid coding for a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide. In such cells, the nucleic acid may be under the control of its genomic regulatory elements, or may be under the control of exogenous regulatory elements including an exogenous promoter. By "exogenous" it is meant a promoter that is not normally coupled transcriptionally to the coding sequence for the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide in its native state.

The invention features a method for identifying human cells containing a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide or a related sequence. The method involves identifying the novel polypeptide in human cells using techniques that are routine and standard in the art, such as those described herein for identifying PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 (*e.g.*, cloning, Southern or Northern blot analysis, in situ hybridization, PCR amplification, etc.).

The invention also features methods of screening cells for natural binding partners of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptides.

The term "natural binding partner" refers to molecules, or portions of these molecules, that bind to the protein of interest in cells. Natural binding partners may be polypeptides or lipids, but do not include glutathione. Natural binding partners can play a role in propagating a signal in a protein signal transduction process. A change in the interaction between a protein and a natural binding partner can manifest itself as an increased or decreased probability that the interaction forms, or an increased or decreased concentration of the protein/natural binding partner complex.

A protein's natural binding partner can bind to a protein's intracellular region with high affinity. High affinity represents an equilibrium binding constant on the order of 10^{-6} M or less. In addition, a natural binding partner can also transiently interact with a protein's intracellular region and chemically modify it. Natural binding partners of protein are chosen from a group that includes, but is not limited to, *SRC* homology 2 (SH2) or 3 (SH3) domains, other phosphoryl tyrosine binding (PTB) domains, guanine nucleotide exchange factors, protein phosphatases, and other protein kinases or protein phosphatases. Methods of determining changes in interactions between proteins and their natural binding partners are readily available in the art.

In another aspect, the invention provides an assay to identify substances capable of modulating the activity of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7. Such assays may be performed in vitro or in vivo can be obtained by modifying existing assays, such as the assays described in WO 96/40276, published December 19, 1996 and WO 96/14433, published May 17, 1996. Other possibilities include testing for phosphatase

activity on standard substrates such as Src kinase or synthetic amino acid substrates. The substances so identified may be enhances or inhibitors of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity and can be peptides, natural products (such as those isolated from fungal strains, for example) or small molecular weight chemical compounds. A preferred substance will be a compound with a molecular weight of less than 5,000, more preferably less than 1,000, most preferably less than 500. The assay and substances contemplated by the invention are discussed in more detail below.

In a preferred embodiment, the invention provides a method for treating or preventing an abnormal condition by administering a compound which is a modulator of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 function *in vitro*. The abnormal condition preferably involves abnormality in PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 signal transduction pathway, and most preferably is cancer. Such compounds preferably show positive results in one or more *in vitro* assays for an activity corresponding to treatment of the disease or disorder in question (such as the assays described in examples 5, 10, 15, 20, and 21 below). Examples of substances that can be screened for favorable activity are provided in section XIV below.

Substances identified as modulators of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity can be used to study the effects of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 modulation in animal models of cell proliferative disorders. For example, inhibitors of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity can be tested as treatments for cell proliferative disorders such as leukemia or lymphoma using subcutaneous xenograph models in mice.

In a further aspect, the invention provides a method for identifying modulators of protein activity. The method involves the steps of: a) forming a captured protein by

contacting the protein with a natural binding partner; b) contacting the captured protein with a test compound; and c) measuring the protein activity. Preferably, the method also includes the step of comparing the protein activity with the activity of a control protein, which has the same amino acid sequence as the protein in step (a) without the natural binding partner, to determine the extent of modulation.

The term "modulator" refers to a compound which has the ability of altering the activity of a protein. A modulator may activate the activity of the protein, may activate or inhibit the activity of the protein depending on the concentration of the compound exposed to the protein, or may inhibit the activity of the protein.

The term "modulator" also refers to a compound that alters the function of a protein by increasing or decreasing the probability that a complex forms between a protein and a natural binding partner. A modulator preferably increases the probability that such a complex forms between the protein and the natural binding partner, more preferably increases or decreases the probability that a complex forms between the protein and the natural binding partner depending on the concentration of the compound exposed to the protein, and most preferably decreases the probability that a complex forms between the protein and the natural binding partner.

The term "activity of a protein", in the context of the invention, defines the natural function of a protein in a cell. Examples of protein function include, but are not limited to, catalytic activity and binding a natural binding partner.

The term "activates" refers to increasing the natural function of a protein. The protein function is preferably the interaction with a natural binding partner and most preferably catalytic activity.

The term "inhibit" refers to decreasing the cellular function of a protein. The protein function is preferably the interaction with a natural binding partner and most preferably catalytic activity.

The term "catalytic activity", in the context of the invention, defines the rate at which a protein reacts with a substrate. Catalytic activity can be measured, for example, by determining the amount of a substrate converted to a product as a function of time. When the protein is a protein kinase or a protein phosphatase, then the reaction with a substrate is the phosphorylation or dephosphorylation of the substrate, respectively. Phosphorylation or dephosphorylation of a substrate occurs at the active-site of a protein kinase or a protein phosphatase. The active-site is normally a cavity in which the substrate binds to the protein kinase or protein phosphatase and is phosphorylated.

The term "substrate" as used herein refers to a molecule which is acted upon by an enzyme. If the enzyme is a protein kinase then the substrate is phosphorylated by the protein kinase. If the enzyme is a protein phosphatase then the substrate is dephosphorylated by the protein phosphatase.

The term "compound" refers to a molecule which has at least two types of atoms in its composition. The molecule may be a small organic molecule. The term "organic molecule" refers to a molecule which has at least one carbon atom in its structure.

The term "complex" refers to an assembly of at least two molecules bound to one another. Signal transduction complexes often contain at least two protein molecules bound to one another. For instance, a protein tyrosine receptor protein kinase, GRB2, SOS, RAF, and RAS assemble to form a signal transduction complex in response to a mitogenic ligand.

The term "contacting" as used herein refers to any touching between a compound and a protein, preferably the mixing of a

solution comprising a compound with a liquid medium bathing the protein of the methods. The touching may involve interaction between the compound and the protein. The solution comprising the compound may be added to the medium bathing the protein by utilizing a delivery apparatus, such as a pipet-based device or syringe-based device.

The term "protein" as used herein refers to a naturally occurring or chemically modified polypeptide chain that has distinct secondary and tertiary structures. The chemical modification may be point mutations. The term "protein" as used herein does not include a polypeptide chain which is covalently fused or otherwise joined through human intervention with another distinct polypeptide chain. For example, a GST-fusion protein is not included under the term "protein" as used herein.

The term "captured protein" as used herein refers to a protein that has come to contact with one of its natural binding partners and has formed a complex with the natural binding partner. The natural binding partner may be free in the solution, bound to a solid support, or free in the solution with the ability to bind to a solid support.

The term "test compound" refers to a compound under study for its potential effect on the catalytic activity of a protein.

The term "control protein" refers to a protein which has the same amino acid sequence of the captured protein but is not being modulated by a test compound, nor has it come in contact with a test compound, nor is it bound to a natural binding partner. The activity of a control protein can be measured using the techniques of the invention, and such activity may be compared with the activity of a modulated protein. A difference between the levels of the two measured activities determines the extent of modulation by the modulators.

The invention provides a method for identifying modulators of protein activity, where the method is preferably a non-radioactive method. The protein is preferably not a fusion protein. Most preferably, the protein is not a GST-fusion protein. The protein is preferably an enzyme, a receptor enzyme, or a non-receptor enzyme, more preferably a protein kinase, and most preferably a protein tyrosine kinase. The protein tyrosine kinase is preferably Zap70 or Syk. In other preferred embodiments, the protein is a protein tyrosine phosphatase, and more preferably the protein is PTP04, SAD, PTP05, PTP10, ALP, or ALK-7.

The term "fusion protein" refers to a heterologous protein formed by the covalent linkage of two distinct polypeptides. The term "GST-fusion protein" refers to a heterologous protein formed by the covalent linkage of a polypeptide and glutathione S-transferase (GST).

The term "enzyme" refers to a protein that can act as a catalyst for biological reactions. Examples of catalyzed biological reactions include, but are not limited to, formation of new bonds, addition of water, addition of a phosphoryl group, and isomerization of an organic molecule.

The term "catalyst" refers to a compound or a dissolved metal ion that increases the rate of a chemical reaction without being consumed in the reaction.

The term "receptor enzyme" refers to an enzyme that has a portion of its amino acid sequence within the cell membrane.

The term "non-receptor enzyme" refers to an enzyme that has none of its amino acid sequence within the cell membrane. The non-receptor enzyme may be associated with the membrane via interactions, such as covalent linkage with fatty acids of the membrane.

The term "protein kinase" refers to an enzyme that transfers the high energy phosphate of adenosine triphosphate

to an amino acid residue, either tyrosine, serine, or threonine, located on a protein target.

The term "protein tyrosine kinase," or PTK, refers to an enzyme that transfers the high energy phosphate of adenosine triphosphate to a tyrosine residue located on a protein target.

"Zap70" and "Syk" are protein tyrosine kinases of the Syk family which is characterized by the presence of two tandemly arranged Src-homology 2 (SH2) domains and no membrane localization motifs. These proteins are probably phosphorylated by the Src family of protein tyrosine kinases at the two tyrosine residues within the ITAM motif.

The term "ITAM motif" stands for "immunoreceptor tyrosine-based activation motif" and refers to a 16 amino acid motif ($YXXLX_{6-8}YXXL$) that is conserved in all of the signal transducing subunits of the T-cell antigen receptor (TCR) (*c.f.* Chan, et al. (1995) *The EMBO Journal*, 14:11, 2499-2508).

The term "protein tyrosine phosphatase" refers to an enzyme that removes a phosphate group from a phosphotyrosine in a protein target.

In a preferred embodiment, the natural binding partner of one of the above proteins is capable of binding to a solid support. The natural binding partner is preferably a peptide, more preferably a phosphopeptide, and most preferably the phosphopeptide comprises an ITAM motif. In other preferred embodiments, the natural binding partner comprises a lipid.

The term "solid support" as used herein refers to an insoluble surface to which a molecule can be bound. Examples of solid supports include, but are not limited to, well plates (*i.e.* 96-well plates), glass beads, or resins (*i.e.* cellulose, agarose, polypropylene, polystyrene, etc.). Natural binding partners can be attached, through either covalent or non-covalent interactions, to the solid support prior to or after binding a protein. Examples of non-covalent interactions

include, but are not limited to, hydrogen bonds, electrostatic interactions, and hydrophobic interactions.

The term "peptide" refers to an arrangement of two or more amino acids, linked together through an amide bond between the carboxyl end of one amino acid and the amino end of another.

The term "phosphopeptide" refers to a peptide that has a phosphate group chemically attached to one of its amino acid residues.

The term "lipid" refers to a water-insoluble substance that can be extracted from cells by organic solvents of low polarity. Examples of lipids include, but are not limited to, glycerides, steroids, and terpenes.

The modulators of protein activity being identified by the methods of the invention preferably modulate the autocatalytic activity, catalytic activity, or binding of a second natural binding partner.

The activity of an enzyme is "autocatalytic activity" when the enzyme and its substrate are identical. Some receptor protein tyrosine kinases are capable of exhibiting autocatalytic activity.

In preferred embodiments, the invention provides a method for identifying modulators of protein activity, comprising the step of contacting the captured protein with one or more components of the group consisting of a substrate, a second natural binding partner, and an antibody. The method preferably further involves the step of lysing cells before forming the captured protein. Most preferably, the method involves the step of washing the solid support after capturing the protein and binding the protein:natural binding partner complex to the solid support and prior to measuring the protein activity.

In another aspect, the invention provides a kit for the identification of modulators of non-receptor enzyme activity

comprising: a) a natural binding partner; b) a solid support; and c) one or more components selected from the group consisting of a substrate, a second natural binding partner, and an antibody.

The natural binding partner in the above kit is preferably a peptide, more preferably a phosphopeptide. Even more preferably the phosphopeptide comprises an ITAM motif. In other preferred embodiment, the natural binding partner comprises a lipid.

The summary of the invention described above is non-limiting and other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Figures

Figure 1 shows a comparison between the amino acid sequence of human PTP04 and the amino acid sequence of the protein to which it is most closely related, murine ZPEP. The relative homology between the two (approximately 70%) suggests that the two proteins are members of the same PTP family but are not species orthologs.

Detailed Description of the Invention

The present invention relates to the isolation and characterization of new proteins which we have called PTP04, SAD, PTP05, PTP10, ALP, and ALK-7, nucleotide sequences encoding PTP04, SAD, PTP05, PTP10, ALP, or ALK-7, various products and assay methods that can be used to identify compounds useful for the diagnosis and treatment of various PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 related diseases and conditions, for example cancer. Polypeptides derived from PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 and nucleic acids encoding such polypeptides may be produced using well known and

standard synthesis techniques when given the sequences presented herein.

I. The Polypeptides of the Invention

A. PTP04

PTP04 is a tyrosine phosphatase with an apparent molecular weight of approximately 100 kDa. Primary sequence analysis shows that PTP04 is comprised of three domains: an N-terminal domain, a catalytic domain, and a C-terminal domain. The lack of a hydrophobic stretch of amino acids generally characterized as a transmembrane region indicates that PTP04 is a non-receptor tyrosine phosphatase.

The full-length PTP04 was originally isolated from a human leukemia cell line. Subsequent expression analysis of both normal tissues and cancer cell lines, shown in detail below, revealed that PTP04 is expressed in human thymus and has very low expression in other normal cells but is significantly overexpressed in a number of tumors, particularly in leukemias and lymphomas. This suggests that PTP04 plays an important role in the growth and persistence of these cancers.

B. SAD

SAD is a tyrosine kinase with an apparent molecular weight of approximately 55 kDa. Primary sequence analysis shows that SAD is comprised of four domains: a domain at the N-terminus that shows no homology to any known sequence (the unique domain), an SH3 domain, an SH2 domain and a catalytic domain. The lack of a hydrophobic stretch of amino acids generally characterized as a transmembrane region indicates that SAD is a non-receptor tyrosine kinase. A comparison of the amino acid sequences suggests that SAD is a member of the Frk family. Like some other members of this family, SAD lacks an N-terminal

myristylation site and a C-terminal regulatory tyrosine characteristic of Src family members. It is most closely related to the murine NR-TK Srm (Kohmura, et al, Mol. Cell. Bio. 14(10):6915, 1994) with approximately 85% sequence homology in the catalytic domain. (Discussed in detail in the examples below.)

SAD was originally isolated from a human breast cancer cell line. Subsequent expression analysis of both normal tissues and cancer cell lines, shown in detail below, revealed that SAD has very limited expression in normal cells but is significantly overexpressed in a number of tumors. This suggests that SAD plays an important role in the growth and persistence of these cancers.

C. PTP05 and PTP10

PTP05 is a tyrosine phosphatase with an apparent molecular weight of approximately 49 kDa. Two additional isoforms have been identified, one larger (approximately 54 kDa) and one smaller (approximately 47 kDa). Primary sequence analysis shows that PTP05 is comprised of three domains: an N-terminal domain, a catalytic domain, and a C-terminal domain. The lack of a hydrophobic stretch of amino acids generally characterized as a transmembrane region indicates that PTP05 is a non-receptor tyrosine phosphatase. PTP10 is also a tyrosine phosphatase with significant homology to PTP05. Together they define a new family of PTPs.

D. ALP

ALP is a tyrosine phosphatase with an apparent molecular weight of approximately 160 - 200 kDa. Primary sequence analysis shows that ALP is comprised of three domains: a domain at the N-terminus that is rich in proline residues (30.6%) and contains several tyrosines that may be

phosphorylated, a catalytic domain, and a C-terminal domain containing region rich in prolines and serines (45.6%) that resenbling a PEST motif (Rogers, et al, Science 234:364, 1986). These proline rich regions may be protein:protein interaction sites as SH3 domains have been shown to bind to proline rich regions (Morton and Campbell, Curr. Biol. 4:614, 1994; Ren, et al, Science 259:1157, 1993). The lack of a hydrophobic stretch of amino acids generally characterized as a transmembrane region indicates that ALP is a non-receptor tyrosine phosphatase.

The full-length ALP was originally isolated from a human brain cancer cell line. Subsequent expression analysis of both normal tissues and cancer cell lines, shown in detail below, revealed that ALP has low expression in normal cells but is significantly overexpressed in a number of tumors. This suggests that ALP plays an important role in the growth and persistence of these cancers.

E. ALK-7

ALK-7 is a type I receptor serine/threonine kinase (STK receptor). Proteins with some homology have been described in the rat (Ryden, et al. J. Biol. Chem. 271:30603, 1996; Tsuchida, et al. Molec. Cell. Neurosci. 7:467, 1996), however, unlike the rat proteins, the human ALK-7 is expressed in more restricted regions of the brain, notably hippocampous, hypothalamic nuclei, sustantia nigra, an pituitary. This extremely restricted expression pattern strongly suggests a role for human ALK-7 in the growth and/or survival of neurons and its relevance in treatment of such diseases as Parkinson's, Huntington's disease and Alzheimer's.

The polypeptide and nucleotide sequences of the invention can be used, therefore, to identify modulators of cell growth

and survival which are useful in developing therapeutics for various cell proliferative disorders and conditions, and in particular cancers related to inappropriate PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity. Assays to identify compounds that act intracellularly to enhance or inhibit PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity can be developed by creating genetically engineered cell lines that express PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 nucleotide sequences, as is more fully discussed below.

II. Nucleic Acids Encoding the Polypeptides of the Invention.

A first aspect of the invention features nucleic acid sequences encoding a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide. Included within the scope of this invention are the functional equivalents of the herein-described isolated nucleic acid molecules. Functional equivalents or derivatives can be obtained in several ways. The degeneracy of the genetic code permits substitution of certain codons by other codons which specify the same amino acid and hence would give rise to the same protein. The nucleic acid sequence can vary substantially since, with the exception of methionine and tryptophan, the known amino acids can be coded for by more than one codon. Thus, portions or all of the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene could be synthesized to give a nucleic acid sequence significantly different from that shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8. The encoded amino acid sequence thereof would, however, be preserved.

In addition, the nucleic acid sequence may comprise a nucleotide sequence which results from the addition, deletion or substitution of at least one nucleotide to the 5'-end and/or the 3'-end of the nucleic acid formula shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID

NO:6, SEQ ID NO:7, or SEQ ID NO:8, or a derivative thereof. Any nucleotide or polynucleotide may be used in this regard, provided that its addition, deletion or substitution does not alter the amino acid sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, or SEQ ID NO:16 which is encoded by the nucleotide sequence. For example, the present invention is intended to include any nucleic acid sequence resulting from the addition of ATG as an initiation codon at the 5'-end of the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 nucleic acid sequence or its functional derivative, or from the addition of TTA, TAG or TGA as a termination codon at the 3'-end of the inventive nucleotide sequence or its derivative. Moreover, the nucleic acid molecule of the present invention may, as necessary, have restriction endonuclease recognition sites added to its 5'-end and/or 3'-end.

Such functional alterations of a given nucleic acid sequence afford an opportunity to promote secretion and/or processing of heterologous proteins encoded by foreign nucleic acid sequences fused thereto. All variations of the nucleotide sequence of the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 genes and fragments thereof permitted by the genetic code are, therefore, included in this invention.

Further, it is possible to delete codons or to substitute one or more codons by codons other than degenerate codons to produce a structurally modified polypeptide, but one which has substantially the same utility or activity of the polypeptide produced by the unmodified nucleic acid molecule. As recognized in the art, the two polypeptides are functionally equivalent, as are the two nucleic acid molecules which give rise to their production, even though the differences between the nucleic acid molecules are not related to degeneracy of the genetic code.

Functional equivalents or derivatives of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 can also be obtained using nucleic acid molecules encoding one or more functional domains of the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide.

The catalytic domain of PTP04 functions as an enzymatic remover of phosphate molecules bound onto tyrosine amino acids and a nucleic acid sequence encoding the catalytic domain alone or linked to other heterologous nucleic acid sequences can be considered a functional derivative of PTP04. Other functional domains of PTP04 include, but are not limited to, the proline-rich region within the N-terminal domain, and the C-terminal domain. Nucleic acid sequences encoding these domains are shown in SEQ ID NO:1 as follows: N-terminal domain 53-196; catalytic domain 197-934, C-terminal domain 935-2473.

The SH2 domain of SAD functions as a phosphorylated tyrosine binding domain and a nucleic acid sequence encoding the SH2 domain alone or linked to other heterologous nucleic acid sequences can be considered a functional derivative of SAD. Other functional domains of SAD include, but are not limited to, the unique domain, the SH3 domain, and the catalytic domain. Nucleic acid sequences encoding these domains are shown in SEQ ID NO:2 as follows: N-terminal unique domain approximately 49-213; SH3 domain approximately 214-375; SH2 domain approximately 406-684; catalytic domain approximately 736-1488.

The catalytic domain of PTP05 functions to remove phosphate molecules bound onto tyrosine residues and a nucleic acid sequence encoding the catalytic domain alone or linked to other heterologous nucleic acid sequences can be considered a functional derivative of PTP05. Other functional domains of these proteins include, but are not limited to, the proline-rich region within the N-terminal domain, and the C-terminal domain. Nucleic acid sequences encoding these domains are

shown in SEQ ID NO:3 as follows: N-terminal domain approximately 199-759 ; catalytic domain approximately 760-1458, C-terminal domain approximately 1459-1476.

The N-terminal proline-rich domain of ALP functions as a SH3 binding domain and a nucleic acid sequence encoding the N-terminal proline-rich domain alone or linked to other heterologous nucleic acid sequences can be considered a functional derivative of ALP. Other functional domains of ALP include, but are not limited to, the proline-rich region within the N-terminal proline-rich domain, the C-terminal proline/serine-rich domain, the proline/serine-rich region within the C-terminal proline/serin-rich domain, and the catalytic domain. Nucleic acid sequences encoding these domains are shown in SEQ ID NO:7 as follows: N-terminal domain 313-2883; proline-rich region 1369-2643 ; catalytic domain approximately 2884-3600, C-terminal proline/serine-rich domain 3601-4134, proline/serine-rich region 3613-4456.

The extracellular domain of ALK-7 functions as a ligand or co-receptor binding domain and a nucleic acid sequence encoding the extracellular domain alone or linked to other heterologous nuclic acid sequences can be considered a functional derivative of ALK-7. Other functional domains of ALK-7 include, but are not limited to, the signal sequence, the transmembrane domain, the intracellular domain, and the catalytic domain. Nucleic acid sequences encoding these domains are shown in SEQ ID NO:8 as follows: signal sequence 155-229; extracellular domain 155-493; transmembrane domain 494-568; intracellular domain 569-1633; catalytic domain approximately 731-1609. It should be noted that the signal sequence is cleaved from the extracellular domain in the mature protein.

III. A Nucleic Acid Probe for the Detection of the Proteins of the Invention.

A nucleic acid probe of the present invention may be used to probe an appropriate chromosomal or cDNA library by usual hybridization methods to obtain another nucleic acid molecule of the present invention. A chromosomal DNA or cDNA library may be prepared from appropriate cells according to recognized methods in the art (*e.g.* "Molecular Cloning: A Laboratory Manual", second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989).

In the alternative, chemical synthesis is carried out in order to obtain nucleic acid probes having nucleotide sequences which correspond to N-terminal and C-terminal portions of the amino acid sequence of the polypeptide of interest. Thus, the synthesized nucleic acid probes may be used as primers in a polymerase chain reaction (PCR) carried out in accordance with recognized PCR techniques, essentially according to PCR Protocols, "A Guide to Methods and Applications", edited by Michael et al., Academic Press, 1990, utilizing the appropriate chromosomal or cDNA library to obtain the fragment of the present invention.

One skilled in the art can readily design such probes based on the sequence disclosed herein using methods of computer alignment and sequence analysis known in the art (*e.g.*. "Molecular Cloning: A Laboratory Manual", second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, 1989). The hybridization probes of the present invention can be labeled by standard labeling techniques such as with a radiolabel, enzyme label, fluorescent label, biotin-avidin label, chemiluminescence, and the like. After hybridization, the probes may be visualized using known methods.

The nucleic acid probes of the present invention include RNA as well as DNA probes and nucleic acids modified in the sugar, phosphate or even the base portion as long as the probe still retains the ability to specifically hybridize under conditions as disclosed herein. Such probes are generated using techniques known in the art. The nucleic acid probe may be immobilized on a solid support. Examples of such solid supports include, but are not limited to, plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins, such as polyacrylamide and latex beads, and nitrocellulose. Techniques for coupling nucleic acid probes to such solid supports are well known in the art.

The test samples suitable for nucleic acid probing methods of the present invention include, for example, cells or nucleic acid extracts of cells, or biological fluids. The sample used in the above-described methods will vary based on the assay format, the detection method and the nature of the tissues, cells or extracts to be assayed. Methods for preparing nucleic acid extracts of cells are well known in the art and can be readily adapted in order to obtain a sample which is compatible with the method utilized.

IV. A Probe Based Method And Kit For Detecting the Proteins of the Invention.

One method of detecting the presence of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 in a sample comprises (a) contacting the sample with the above-described nucleic acid probe, under conditions such that hybridization occurs, and (b) detecting the presence of the probe bound to the nucleic acid molecule. One skilled in the art would select the nucleic acid probe according to techniques known in the art as described above. Samples to be tested include but should not be limited to RNA samples of human tissue.

A kit for detecting the presence of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 in a sample comprises at least one container having disposed therein the above-described nucleic acid probe. The kit may further comprise other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound nucleic acid probe. Examples of detection reagents include, but are not limited to radiolabelled probes, enzymaticly labeled probes (horseradish peroxidase, Alkaline phosphatase), and affinity labeled probes (biotin, avidin, or steptavidin).

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the probe or primers used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, and the like), and containers which contain the reagents used to detect the hybridized probe, bound antibody, amplified product, or the like. One skilled in the art will readily recognize that the nucleic acid probes described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art with or without a set of instructions concerning the use of such reagents in an assay.

V. DNA Constructs Comprising a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 Nucleic Acid Molecule and Cells Containing These Constructs.

The present invention also relates to a recombinant DNA molecule comprising, 5' to 3', a promoter effective to initiate transcription in a host cell and the above-described nucleic acid molecules. In addition, the present invention relates to a recombinant DNA molecule comprising a vector and a nucleic acid molecule described herein. The present invention also relates to a nucleic acid molecule comprising a transcriptional region functional in a cell, a sequence complimentary to an RNA sequence encoding an amino acid sequence corresponding to a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide or functional derivative, and a transcriptional termination region functional in said cell. The above-described molecules may be isolated and/or purified DNA molecules.

The present invention also relates to a cell or organism that contains a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 nucleic acid molecule as described herein and thereby is capable of expressing a peptide. The polypeptide may be purified from cells which have been altered to express the polypeptide. A cell is said to be "altered to express a desired polypeptide" when the cell, through genetic manipulation, is made to produce a protein which it normally does not produce or which the cell normally produces at lower levels. One skilled in the art can readily adapt procedures for introducing and expressing either genomic, cDNA, or synthetic sequences into either eukaryotic or prokaryotic cells.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains nucleotide sequences which contain transcriptional and translational regulatory information and such sequences are "operably linked" to nucleotide sequences which encode the polypeptide. An

operable linkage is a linkage in which the regulatory DNA sequences and the DNA sequence sought to be expressed are connected in such a way as to permit gene sequence expression. The precise nature of the regulatory regions needed for gene sequence expression may vary from organism to organism, but will in general include a promoter region which, in prokaryotes, contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like.

If desired, the non-coding region 3' to the sequence encoding a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene may be obtained by the above-described cloning methods. This region may be retained for its transcriptional termination regulatory sequences, such as termination and polyadenylation. Thus, by retaining the 3'-region naturally contiguous to the DNA sequence encoding a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene, the transcriptional termination signals may be provided. Where the transcriptional termination signals are not satisfactorily functional in the expression host cell, then a 3' region functional in the host cell may be substituted.

Two DNA sequences (such as a promoter region sequence and a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 sequence) are said to be operably linked if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region sequence to direct the transcription of a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene sequence, or (3) interfere with the ability of the a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene sequence to be transcribed by the promoter

region sequence. Thus, a promoter region would be operably linked to a DNA sequence if the promoter were capable of effecting transcription of that DNA sequence. Thus, to express a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene, transcriptional and translational signals recognized by an appropriate host are necessary.

The present invention encompasses the expression of a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene (or a functional derivative thereof) in either prokaryotic or eukaryotic cells. Prokaryotic hosts are, generally, very efficient and convenient for the production of recombinant proteins and are, therefore, one type of preferred expression system for a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene. Prokaryotes most frequently are represented by various strains of E. coli. However, other microbial strains may also be used, including other bacterial strains.

In prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host may be used. Examples of suitable plasmid vectors may include pBR322, pUC118, pUC119 and the like; suitable phage or bacteriophage vectors may include lgt10, lgt11 and the like; and suitable virus vectors may include pMAM-neo, pKRC and the like. Preferably, the selected vector of the present invention has the capacity to replicate in the selected host cell.

Recognized prokaryotic hosts include bacteria such as E. coli and those from genera such as Bacillus, Streptomyces, Pseudomonas, Salmonella, Serratia, and the like. However, under such conditions, the polypeptide will not be glycosylated. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

To express PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 (or a functional derivative thereof) in a prokaryotic cell, it is

necessary to operably link a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 sequence to a functional prokaryotic promoter. Such promoters may be either constitutive or, more preferably, regulatable (*i.e.*, inducible or derepressible). Examples of constitutive promoters include the int promoter of bacteriophage l, the bla promoter of the b-lactamase gene sequence of pBR322, and the CAT promoter of the chloramphenicol acetyl transferase gene sequence of pPR325, and the like. Examples of inducible prokaryotic promoters include the major right and left promoters of bacteriophage l (P_L and P_R), the trp, recA, lacZ, lacI, and gal promoters of E. coli, the a-amylase (Ulmanen et at., J. Bacteriol. 162:176-182, 1985) and the sigma-28-specific promoters of B. subtilis (Gilman et al., Gene sequence 32:11-20(1984)), the promoters of the bacteriophages of Bacillus (Gryczan, In: The Molecular Biology of the Bacilli, Academic Press, Inc., NY (1982)), and Streptomyces promoters (Ward et at., Mol. Gen. Genet. 203:468-478, 1986). Prokaryotic promoters are reviewed by Glick (J. Ind. Microbiot. 1:277-282, 1987); Cenatiempo (Biochimie *68*:505-516, 1986); and Gottesman (Ann. Rev. Genet. 18:415-442, 1984).

Proper expression in a prokaryotic cell also requires the presence of a ribosome binding site upstream of the gene sequence-encoding sequence. Such ribosome binding sites are disclosed, for example, by Gold et at. (Ann. Rev. Microbiol. 35:365-404, 1981). The selection of control sequences, expression vectors, transformation methods, and the like, are dependent on the type of host cell used to express the gene.

As used herein, "cell", "cell line", and "cell culture" may be used interchangeably and all such designations include progeny. Thus, the words "transformants" or "transformed cells" include the primary subject cell and cultures derived therefrom, without regard to the number of transfers. It is also understood that all progeny may not be precisely identical

in DNA content, due to deliberate or inadvertent mutations. However, as defined, mutant progeny have the same functionality as that of the originally transformed cell.

Host cells which may be used in the expression systems of the present invention are not strictly limited, provided that they are suitable for use in the expression of the PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 peptide of interest. Suitable hosts may often include eukaryotic cells. Preferred eukaryotic hosts include, for example, yeast, fungi, insect cells, mammalian cells either in vivo, or in tissue culture. Mammalian cells which may be useful as hosts include HeLa cells, cells of fibroblast origin such as VERO, 3T3 or CHO-K1, or cells of lymphoid origin (such as 32D cells) and their derivatives. Preferred mammalian host cells include SP2/0 and J558L, as well as neuroblastoma cell lines such as IMR 332 and PC12 which may provide better capacities for correct post-translational processing.

In addition, plant cells are also available as hosts, and control sequences compatible with plant cells are available, such as the cauliflower mosaic virus 35S and 19S, and nopaline synthase promoter and polyadenylation signal sequences. Another preferred host is an insect cell, for example the Drosophila larvae. Using insect cells as hosts, the Drosophila alcohol dehydrogenase promoter can be used. Rubin, Science 240:1453-1459, 1988). Alternatively, baculovirus vectors can be engineered to express large amounts of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 in insects cells (Jasny, Science 238:1653, 1987); Miller et al., In: Genetic Engineering (1986), Setlow, J.K., et al., eds., Plenum, Vol. 8, pp. 277-297).

Any of a series of yeast gene sequence expression systems can be utilized which incorporate promoter and termination elements from the actively expressed gene sequences coding for glycolytic enzymes are produced in large quantities when yeast

are grown in mediums rich in glucose. Known glycolytic gene sequences can also provide very efficient transcriptional control signals. Yeast provides substantial advantages in that it can also carry out post-translational peptide modifications. A number of recombinant DNA strategies exist which utilize strong promoter sequences and high copy number of plasmids which can be utilized for production of the desired proteins in yeast. Yeast recognizes leader sequences on cloned mammalian gene sequence products and secretes peptides bearing leader sequences (*i.e.*, pre-peptides). For a mammalian host, several possible vector systems are available for the expression of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7.

A particularly preferred yeast expression system is that utilizing Schizosaccharmocyces pombe. This system is useful for studying the activity of members of the Src family (Superti-Furga, et al, EMBO J. 12:2625, 1993) and other NR-TKs.

A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals may be derived from viral sources, such as adenovirus, bovine papilloma virus, cytomegalovirus, simian virus, or the like, where the regulatory signals are associated with a particular gene sequence which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, and the like, may be employed. Transcriptional initiation regulatory signals may be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical (such as metabolite) regulation.

Expression of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 in eukaryotic hosts requires the use of eukaryotic regulatory regions. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis. Preferred eukaryotic promoters include, for example, the promoter of the mouse metallothionein I gene sequence (Hamer et al., J. Mol. Appl. Gen. 1:273-288, 1982); the TK promoter of Herpes virus (McKnight, Cell 31:355-365, 1982); the SV40 early promoter (Benoist et al., Nature (London) 290:304-310, 1981); the yeast gal4 gene sequence promoter (Johnston et al., Proc. Natl. Acad. Sci. (USA) 79:6971-6975, 1982); Silver et al., Proc. Natl. Acad. Sci. (USA) 81:5951-5955, 1984).

Translation of eukaryotic mRNA is initiated at the codon which encodes the first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a DNA sequence which encodes PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 (or a functional derivative thereof) does not contain any intervening codons which are capable of encoding a methionine (*i.e.*, AUG). The presence of such codons results either in a formation of a fusion protein (if the AUG codon is in the same reading frame as a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 coding sequence) or a frame-shift mutation (if the AUG codon is not in the same reading frame as a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 coding sequence).

A PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 nucleic acid molecule and an operably linked promoter may be introduced into a recipient prokaryotic or eukaryotic cell either as a nonreplicating DNA (or RNA) molecule, which may either be a linear molecule or, more preferably, a closed covalent circular molecule (a plasmid). Since such molecules are incapable of autonomous replication, the expression of the gene may occur through the transient expression of the introduced sequence. Alternatively, permanent or stable expression may occur through

the integration of the introduced DNA sequence into the host chromosome.

A vector may be employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may provide for prototrophy to an auxotrophic host, biocide resistance, *e.g.*, antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene sequence can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection. Additional elements may also be needed for optimal synthesis of single chain binding protein mRNA. These elements may include splice signals, as well as transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama, Mol. Cell. Bio. 3:280, 1983.

The introduced nucleic acid molecule can be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose. Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Preferred prokaryotic vectors include plasmids such as those capable of replication in E. coil (such as, for example, pBR322, ColE1, pSC101, pACYC 184, pVX. Such plasmids are, for example, disclosed by Sambrook (cf. "Molecular Cloning: A

Laboratory Manual", second edition, edited by Sambrook, Fritsch, & Maniatis, Cold Spring Harbor Laboratory, (1989)). Bacillus plasmids include pC194, pC221, pT127, and the like. Such plasmids are disclosed by Gryczan (In: The Molecular Biology of the Bacilli, Academic Press, NY (1982), pp. 307-329). Suitable Streptomyces plasmids include p1J101 (Kendall et al., J. Bacteriol. 169:4177-4183,1987), and streptomyces bacteriophages such as fC31 (Chater et al., In: Sixth International Symposium on Actinomycetales Biology, Akademiai Kaido, Budapest, Hungary (1986), pp. 45-54). Pseudomonas plasmids are reviewed by John et al. (Rev. Infect. Dis. 8:693-704, 1986), and Izaki (Jpn. J. Bacteriol. 33:729-742, 1978).

Preferred eukaryotic plasmids include, for example, BPV, vaccinia, SV40, 2-micron circle, and the like, or their derivatives. Such plasmids are well known in the art (Botstein et al., Miami Wntr. Symp. 19:265-274, 1982); Broach, In: "The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance", Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 445-470 (1981); Broach, Cell 28:203-204, 1982); Bollon et at., J. Clin. Hematol. Oncol. 10:39-48, 1980); Maniatis, In: Cell Biology: A Comprehensive Treatise, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563-608 (1980).

Once the vector or nucleic acid molecule containing the construct(s) has been prepared for expression, the DNA construct(s) may be introduced into an appropriate host cell by any of a variety of suitable means, *i.e.*, transformation, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate-precipitation, direct microinjection, and the like. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. Expression of the cloned gene molecule(s) results in the

production of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 or fragments or functional derivatives thereof. This can take place in the transformed cells as such, or following the induction of these cells to differentiate (for example, by administration of bromodeoxyuracil to neuroblastoma cells or the like). A variety of incubation conditions can be used to form the peptide of the present invention. The most preferred conditions are those which mimic physiological conditions.

VI. The Polypeptides of the Invention.

Also a feature of the invention are PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptides. A variety of methodologies known in the art can be utilized to obtain the polypeptides of the present invention. They may be purified from tissues or cells which naturally produce them. Alternatively, the above-described isolated nucleic acid sequences can be used to express a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 protein recombinantly.

Any eukaryotic organism can be used as a source for the polypeptide of the invention, as long as the source organism naturally contains such a polypeptide. As used herein, "source organism" refers to the original organism from which the amino acid sequence is derived, regardless of the organism the protein is expressed in and ultimately isolated from.

One skilled in the art can readily follow known methods for isolating proteins in order to obtain the peptide free of natural contaminants. These include, but are not limited to: size-exclusion chromatography, HPLC, ion-exchange chromatography, and immuno-affinity chromatography.

A PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 protein, like all proteins, is comprised of distinct functional units or domains. In eukaryotes, proteins sorted through the so-called vesicular pathway (bulk flow) usually have a signal sequence

(also called a leader peptide) in the N- terminus, which is cleaved off after the translocation through the ER (endoplasmic reticulum) membrane. Some N-terminal signal sequences are not cleaved off, remaining as transmembrane segments, but it does not mean these proteins are retained in the ER; they can be further sorted and included in vesicles.

SAD protein lacks a hydrophobic signal sequence and is classified as a non-receptor protein. Other motifs involved in targeting proteins to specific cellular locations include those selective for the mitochondrial matrix (Gavel and von Heijne, Prot Eng 4:33, 1990), the nucleus (Robbins, et al, Cell 64:615, 1991), peroxisomes, endoplasmic reticulum (Jackson, et al, EMBO J 9:3253, 1990), vesicular pathways (Bendiak, Biophys Res Comm 170:879, 1990), glycosyl-phosphatidylinositol (GPI) lipid anchors, and lysosomal organelles, and motifs that target proteins to lipid membranes such as myristylation (Towler, et al, Annu Rev Biochem 57:69, 1988) and farnesylation sites. The N-terminal 15 amino acids of the SAD protein conforms to the features which define a mitochondrial membrane protein with a bipartite structure of an N-terminal stretch of high arginine content involved in membrane targeting followed by the apolar sequence which signals translocation to the mitochondrial intermembrane space.

Non-receptor proteins generally function to transmit signals within the cell, either by providing sites for protein:protein interactions or by having some catalytic activity (contained within a catalytic domain), often both. Methods of predicting the existence of these various domains are well known in the art. Protein:protein interaction domains can be identified by comparison to other proteins. The SH2 domain, for example is a protein domain of about 100 amino acids first identified as a conserved sequence region between the proteins Src and Fps (Sadowski, et al, Mol. Cell. Bio.

6:4396, 1986). Similar sequences were later found in many other intracellular signal-transducing proteins. SH2 domains function as regulatory modules of intracellular signaling cascades by interacting with high affinity to phosphotyrosine-containing proteins in a sequence specific and strictly phosphorylation-dependent manner (Mayer and Baltimore, Trends Cell. Biol. 3:8, 1993). Kinase or phosphatase catalytic domains can be identified by comparison to other known catalytic domains with kinase or phosphatase activity. See, for example Hanks and Hunter, FASEB J. 9:576-595, 1995.

Receptor proteins also have, and are somewhat defined by, a hydrophobic transmembrane segment(s) which are thought to be Alpha-helices in membranes. Membrane proteins also integrate into the cell membrane in a specific manner with respect to the two sides (cytoplasmic/intracellular or exo-cytoplasmic/ extracellular), which is referred to as membrane topology. Extracellular portions of integral membrane proteins often function as ligand binding domains whereas intracellula portions generally function to transmit signals within the cell, either by providing sites for protein:protein interactions or by having some catalytic activity (contained within a catalytic domain), often both. Methods of predicting the existence of these various domains are well known in the art. See, for example, D. J. McGeoch, Virus Research 3:271, 1985, or G. von Heijne, Nucl. Acids Res. 14:4683, 1986, for signal sequences, P. Klein, et al., Biochim. Biophys. Acta 815:468, 1985, for transmembrane domains, and S. J. Singer, Ann. Rev. Cell Biol. 6:247, 1990, or E. Hartmann, et al., Proc. Natl. Acad. Sci. USA, 86:5786, 1989, for prediction of membrane topology. Kinase catalytic domains can be identified by comparison to other known catalytic domains with kinase activity. See, for example, Hanks and Hunter, FASEB J. 9:576-595, 1995.

Primary sequence analysis of the PTP04 amino acid sequence (shown in SEQ ID NO:9) reveals that it does not contain a signal sequence or transmembrane domain and is, therefore, an intracellular protein. Comparison to known protein sequences revels that PTP04 is comprised of several unique domains. These include a 48 amino acid N-terminal domain (shown from amino acid number 1-48 of SEQ ID NO:9), a 245 amino acid catalytic domain (shown from amino acid number 49-294 of SEQ ID NO:9), and a 512 amino acid C-terminal domain (shown from amino acid number 295-807 of SEQ ID NO:9).

Primary sequence analysis of the SAD amino acid sequence (shown in SEQ ID NO:10) reveals that it contains four distinct domains. These include an approximately 55 amino acid N-terminal unique domain (shown from amino acid number 1-55 of SEQ ID NO:10), an approximately 54 amino acid SH3 domain (shown from amino acid number 56-109 of SEQ ID NO:10), an approximately 93 amino acid SH2 domain (shown from amino acid number 120-212 of SEQ ID NO:10), an approximately 251 amino acid catalytic domain (amino acid number 230-480 of SEQ ID No:10), and a C-terminal tail of 8 amino acids (shown from amino acid 481-488 of SEQ ID NO:10).

Primary sequence analysis of the PTP05 amino acid sequence (shown in SEQ ID NO:11 with isoforms shown in SEQ ID NO:12 and SEQ ID NO:13) reveals that it and its isoforms do not contain a signal sequence or transmembrane domain, and it is, therefore, an intracellular protein. Comparison to known protein sequences revels that PTP05 is comprised of several unique domains. These include a 187 amino acid N-terminal domain (shown from amino acid number 1-187 of SEQ ID NO:11), a 242 amino acid catalytic domain (shown from amino acid number 188-420 of SEQ ID NO:11), and a 5 amino acid C-terminal domain (shown from amino acid number 421-426 of SEQ ID NO:11).

Two additional isoforms of PTP05 were also identified, a "long" form (SEQ ID NO:12) and a "C-trunc" form (SEQ ID NO:13). The "long" form has a 37 amino acid insertion in the N-terminal domain (aminoacids 44-80 of SEQ ID NO:12) which extends this domain to 224 amino acids. The catalytic domain extends from amino acid 225-457 of SEQ ID NO:12 and the C-terminal domain extents from amino acids 458-463 of SEQ ID NO:12. The "C-trunc" form results from a deletion of nucleotides 1415-1507 of SEQ ID NO:3, most likely due to alternative exon splicing. This deletion results in a replacement of the C-terminal 21 amino acids with a unique 7 amino acid sequence. This change eliminates a conserved C-terminal portion of the catalytic domain, which may affect enzymatic activity. The N-terminal domain of the "C-trunc" form extends from amino acid 1-87 of SEQ ID NO:13, the catalytic domain from amino acids 188-405 of SEQ ID NO:13 and the unique C-terminal domain from 406-412 of SEQ ID NO:13.

Primary sequence analysis of the ALP amino acid sequence (shown in SEQ ID NO:15) reveals that it does not contain a signal sequence or transmembrane domain and is, therefore, an intracellular protein. Comparison to known protein sequences revels that ALP is comprised of several unique domains. These include a 857 amino acid N-terminal proline-rich domain (shown from amino acid number 1-857 of SEQ ID NO:15) within which is a proline-rich region (amino acid number 353-777 of SEQ ID NO:15), a 238 amino acid catalytic domain (shown from amino acid number 858-1096 of SEQ ID NO:15), and a 177 amino acid C-terminal proline/serine-rich domain (shown from amino acid number 1097-1274 of SEQ ID NO:15) within which is a proline/serine-rich region (amino acid number 1101-1214 of SEQ ID NO:15).

Primary sequence analysis for an ALK-7 amino acid sequence (shown in SEQ ID NO:16) reveals that it contains all the motifs

characteristic of a type I STK receptor. These include a 25 amino acid signal peptide (shown from amino acid number 1-25 of SEQ ID NO:16), an 88 amino acid cysteine-rich extracellular region (shown from amino acid number 26-113 of SEQ ID NO:16), a single 25 amino acid transmembrane domain (shown from amino acid number 114-136 of SEQ ID NO:16), and a 355 amino acid cytoplasmic domain (shown from amino acid number 137-493 of SEQ ID NO:16), which includes a GS domain and a catalytic domain (amino acid number 193-485 of SEQ ID NO:16).

The extracellular domain conserves the 10 cysteines present in all type I STK receptors (ten Dijke, et al., Oncogene 8:2879, 1993; Bassinge, et al., Science 263:87, 1994; Massague, Trends Cell Biol. 4:172, 1994) and also contains 3 potential N-=linked glycosylation sites. The divergent extracellular domain sequence of ALK-7 (28-30% identity to ALK-4 and ALK-5) suggests it may have a unique ligand/type II STK receptor specificity. A rat ALK-7-like protein ahs been found to bind TGFbeta and activin in a complex with the type II TGF beta receptor and ACTRII. However, these ligands are not expressed in the same cell types as human ALK-7 suggesting alternative ligands. Candidate ALK-7-specific ligands include other TGFbetas such as TGFbeta 2, GDF-1, and homologues of GDNF, such as neuturin, which have been found to be expressed in neurons in a pattern similar to that of ALK-7.

The intracellular domain is somewhat more homologous to other ALK proteins, particularly in the catalytic domain which shows 83% identity to other type I STK receptors. The 40 amino acids immediately N-terminal of the transmembrane domain (the juxtamembrane domain) are, however, quite unique in comparison with other ALKs.

These PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 domains have a variety of uses. An example of such a use is to make a polypeptide consisting of the PTP04, SAD, PTP05, PTP10, ALP, or

ALK-7 catalytic domain and a heterologous protein such as glutathione S-transferase (GST). Such a polypeptide can be used in a biochemical assay for PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 catalytic activity useful for studying PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 substrate specificity or for identifying substances that can modulate PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 catalytic activity. Alternatively, one skilled in the art could create a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide lacking at least one of the three major domains. Such a polypeptide, when expressed in a cell, is able to form complexes with the natural binding partner(s) of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 but unable to transmit any signal further downstream into the cell, *i.e.*,. it would be signaling incompetent and thus would be useful for studying the biological relevance of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity. (See, for example, Gishizky, et al, PNAS :10889, 1995).

VII. An Antibody Having Binding Affinity To the Polypeptides of the Invention And A Hybridoma Containing the Antibody.

The present invention also relates to an antibody having specific binding affinity to an PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide. The polypeptide may have the amino acid sequence set forth in SEQ ID NO:2, or a be fragment thereof, or at least 6 contiguous amino acids thereof. Such an antibody may be identified by comparing its binding affinity to a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide with its binding affinity to another polypeptide. Those which bind selectively to PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 would be chosen for use in methods requiring a distinction between PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 and other polypeptides. Such methods could include, but should not be limited to, the analysis of altered PTP04, SAD, PTP05, PTP10, ALP, or ALK-7

expression in tissue containing other polypeptides and assay systems using whole cells.

A PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 peptide of the present invention can be used to produce antibodies or hybridomas. One skilled in the art will recognize that if an antibody is desired, such a peptide would be generated as described herein and used as an immunogen. Preferred PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 peptides for this purpose as shown in Example 4 below. The antibodies of the present invention include monoclonal and polyclonal antibodies, as well fragments of these antibodies, and humanized forms. Humanized forms of the antibodies of the present invention may be generated using one of the procedures known in the art such as chimerization or CDR grafting. The present invention also relates to a hybridoma which produces the above-described monoclonal antibody, or binding fragment thereof. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing monoclonal antibodies and hybridomas are well known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands, 1984; St. Groth et al., J. Immunol. Methods 35:1-21, 1980). Any animal (mouse, rabbit, and the like) which is known to produce antibodies can be immunized with the selected polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of polypeptide used for immunization will vary based on the animal which is immunized, the antigenicity of the polypeptide and the site of injection.

The polypeptide may be modified or administered in an adjuvant in order to increase the peptide antigenicity.

Methods of increasing the antigenicity of a polypeptide are well known in the art. Such procedures include coupling the antigen with a heterologous protein (such as globulin or b-galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells. Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz, et al., Exp. Cell Res. 175:109-124, 1988). Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", supra, 1984).

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures. The above-described antibodies may be detectably labeled. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, and the like), enzymatic labels (such as horse radish peroxidase, Alkaline phosphatase, and the like) fluorescent labels (such as FITC or rhodamine, and the like), paramagnetic atoms, and the like. Procedures for accomplishing such labeling are well-known in the art, for example, see (Stemberger, et al., J. Histochem. Cytochem. 18:315, 1970; Bayer, et at., Meth. Enzym. 62:308, 1979; Engval, et al., Immunot. 109:129, 1972; Goding, J. Immunol. Meth. 13:215, 1976). The labeled antibodies of the present invention can be

used for in vitro, in vivo, and in in situ assays to identify cells or tissues which express a specific peptide.

The above-described antibodies may also be immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10, 1986; Jacoby et al., Meth. Enzym. 34, Academic Press, N.Y., 1974). The immobilized antibodies of the present invention can be used for in vitro, in vivo, and in situ assays as well as in immunochromotography.

Furthermore, one skilled in the art can readily adapt currently available procedures, as well as the techniques, methods and kits disclosed above with regard to antibodies, to generate peptides capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby et al., "Application of Synthetic Peptides: Antisense Peptides", In Synthetic Peptides, A User's Guide, W.H. Freeman, NY, pp. 289-307(1992), and Kaspczak et al., Biochemistry 28:9230-8(1989).

VIII. An Antibody Based Method And Kit For Detecting the Polypeptides of the Invention.

The present invention encompasses a method of detecting a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide in a sample, comprising: (a) contacting the sample with an above-described antibody, under conditions such that immunocomplexes form, and (b) detecting the presence of said antibody bound to the polypeptide. In detail, the methods comprise incubating a test sample with one or more of the antibodies of the present

invention and assaying whether the antibody binds to the test sample. Altered levels, either an increase or decrease, of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 in a sample as compared to normal levels may indicate disease.

Conditions for incubating an antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the antibody used in the assay. One skilled in the art will recognize that any one of the commonly available immunological assay formats (such as radioimmunoassays, enzyme-linked immunosorbent assays, diffusion based Ouchterlony, or rocket immunofluorescent assays) can readily be adapted to employ the antibodies of the present invention. Examples of such assays can be found in Chard, "An Introduction to Radioimmunoassay and Related Techniques" Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock et al., "Techniques in Immunocytochemistry," Academic Press, Orlando, FL Vol. 1(1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, "Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology," Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The immunological assay test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is capable with the system utilized.

A kit contains all the necessary reagents to carry out the previously described methods of detection. The kit may

comprise: (i) a first container containing an above-described antibody, and (ii) second container containing a conjugate comprising a binding partner of the antibody and a label. In another preferred embodiment, the kit further comprises one or more other containers comprising one or more of the following: wash reagents and reagents capable of detecting the presence of bound antibodies.

Examples of detection reagents include, but are not limited to, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the chromophoric, enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. The compartmentalized kit may be as described above for nucleic acid probe kits. One skilled in the art will readily recognize that the antibodies described in the present invention can readily be incorporated into one of the established kit formats which are well known in the art.

IX. Isolation of Natural Binding Partners of the Polypeptides of the Invention.

The present invention also relates to methods of detecting natural binding partners capable of binding to a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide. A natural binding partner of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 may be, for example, a substrate protein which is dephosphorylated as part of a signaling cascade. The binding partner(s) may be present within a complex mixture, for example, serum, body fluids, or cell extracts.

In general methods for identifying natural binding partners comprise incubating a substance with PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 and detecting the presence of a substance bound to PTP04, SAD, PTP05, PTP10, ALP, or ALK-7.

Preferred methods include the two-hybrid system of Fields and Song (supra) and co-immunoprecipitation.

X. Identification of and Uses for Substances Capable of Modulating the Activity of the Polypeptides of the Invention.

The present invention also relates to a method of detecting a substance capable of modulating PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity. Such substances can either enhance activity (agonists) or inhibit activity (antagonists). Agonists and antagonists can be peptides, antibodies, products from natural sources such as fungal or plant extracts or small molecular weight organic compounds. In general, small molecular weight organic compounds are preferred. Examples of classes of compounds that can be tested for PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 modulating activity are, for example but not limited to, thiazoles (see for example co-pending US applications 60/033,522, 08/660,900), and naphthopyrones (US patent number 5,602,171).

In general the method comprises incubating cells that produce PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 in the presence of a test substance and detecting changes in the level of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity or PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 binding partner activity. A change in activity may be manifested by increased or decreased phosphorylation of a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide, increased or decreased phosphorylation of a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 substrate, or increased or decreased biological response in cells. A method for detecting modulation of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity using the phosphorylation of an artificial substrate is shown in the examples below. Biological responses can include, for example, proliferation, differentiation, survival, or motility.

The substance thus identified would produce a change in activity indicative of the agonist or antagonist nature of the substance. Once the substance is identified it can be isolated using techniques well known in the art, if not already available in a purified form.

The present invention also encompasses a method of agonizing (stimulating) or antagonizing PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 associated activity in a mammal comprising administering to said mammal an agonist or antagonist to PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 in an amount sufficient to effect said agonism or antagonism. Also encompassed in the present application is a method of treating diseases in a mammal with an agonist or antagonist of PTP04-, SAD-, PTP05-, PTP10-, ALP-, or ALK-7-related activity comprising administering the agonist or antagonist to a mammal in an amount sufficient to agonize or antagonize PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 associated function(s). The particular compound can be administered to a patient either by itself or in a pharmaceutical composition where it is mixed with suitable carriers or excipient(s). In treating a patient a therapeutically effective dose of the compound is administered. A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals. For example, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds which exhibit large therapeutic indices are preferred. The data obtained from these cell culture assays

and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized.

For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal disruption of the protein complex, or a half-maximal inhibition of the cellular level and/or activity of a complex component). Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by HPLC.

The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See *e.g.* Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p1).

It should be noted that the attending physician would know how to and when to terminate, interrupt, or adjust administration due to toxicity, or to organ dysfunctions. Conversely, the attending physician would also know to adjust treatment to higher levels if the clinical response were not adequate (precluding toxicity). The magnitude of an administrated dose in the management of the oncogenic disorder of interest will vary with the severity of the condition to be treated and to the route of administration. The severity of the condition may, for example, be evaluated, in part, by standard prognostic evaluation methods. Further, the dose and perhaps dose frequency, will also vary according to the age, body weight, and

response of the individual patient. A program comparable to that discussed above may be used in veterinary medicine.

Depending on the specific conditions being treated, such agents may be formulated and administered systemically or locally. Techniques for formulation and administration may be found in "Remington's Pharmaceutical Sciences," 1990, 18th ed., Mack Publishing Co., Easton, PA. Suitable routes may include oral, rectal, transdermal, vaginal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections, just to name a few.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For such transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

Use of pharmaceutically acceptable carriers to formulate the compounds herein disclosed for the practice of the invention into dosages suitable for systemic administration is within the scope of the invention. With proper choice of carrier and suitable manufacturing practice, the compositions of the present invention, in particular, those formulated as solutions, may be administered parenterally, such as by intravenous injection. The compounds can be formulated readily using pharmaceutically acceptable carriers well known in the art into dosages suitable for oral administration. Such carriers enable the compounds of the invention to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated.

Agents intended to be administered intracellularly may be administered using techniques well known to those of ordinary skill in the art. For example, such agents may be encapsulated into liposomes, then administered as described above. Liposomes are spherical lipid bilayers with aqueous interiors. All molecules present in an aqueous solution at the time of liposome formation are incorporated into the aqueous interior. The liposomal contents are both protected from the external microenvironment and, because liposomes fuse with cell membranes, are efficiently delivered into the cell cytoplasm. Additionally, due to their hydrophobicity, small organic molecules may be directly administered intracellularly.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tablets, dragees, capsules, or solutions.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active com-

pounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added.

XI. Transgenic Animals.

Also contemplated by the invention are transgenic animals useful for the study of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity in complex in vivo systems. A variety of methods are available for the production of transgenic animals associated with this invention. DNA sequences encoding PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 can be injected into the pronucleus of a fertilized egg before fusion of the male and female pronuclei, or injected into the nucleus of an embryonic cell (*e.g.*., the nucleus of a two-cell embryo) following the initiation of cell division (Brinster, et al., Proc. Nat. Acad. Sci. USA *82*: 4438, 1985). Embryos can be infected with viruses, especially retroviruses, modified to carry inorganic-ion receptor nucleotide sequences of the invention.

Pluripotent stem cells derived from the inner cell mass of the embryo and stabilized in culture can be manipulated in culture to incorporate nucleotide sequences of the invention. A transgenic animal can be produced from such cells through implantation into a blastocyst that is implanted into a foster mother and allowed to come to term. Animals suitable for transgenic experiments can be obtained from standard commercial

sources such as Charles River (Wilmington, MA), Taconic (Germantown, NY), Harlan Sprague Dawley (Indianapolis, IN), etc.

The procedures for manipulation of the rodent embryo and for microinjection of DNA into the pronucleus of the zygote are well known to those of ordinary skill in the art (Hogan, et al., supra). Microinjection procedures for fish, amphibian eggs and birds are detailed in Houdebine and Chourrout, Experientia 47: 897-905, 1991). Other procedures for introduction of DNA into tissues of animals are described in U.S. Patent No., 4,945,050 (Sandford et al., July 30, 1990).

By way of example only, to prepare a transgenic mouse, female mice are induced to superovulate. After being allowed to mate, the females are sacrificed by CO_2 asphyxiation or cervical dislocation and embryos are recovered from excised oviducts. Surrounding cumulus cells are removed. Pronuclear embryos are then washed and stored until the time of injection. Randomly cycling adult female mice are paired with vasectomized males. Recipient females are mated at the same time as donor females. Embryos then are transferred surgically. The procedure for generating transgenic rats is similar to that of mice. See Hammer, et al., Cell 63:1099-1112, 1990).

Methods for the culturing of embryonic stem (ES) cells and the subsequent production of transgenic animals by the introduction of DNA into ES cells using methods such as electroporation, calcium phosphate/DNA precipitation and direct injection also are well known to those of ordinary skill in the art. See, for example, Teratocarcinomas and Embryonic Stem Cells, A Practical Approach, E.J. Robertson, ed., IRL Press, 1987).

In cases involving random gene integration, a clone containing the sequence(s) of the invention is co-transfected with a gene encoding resistance. Alternatively, the gene

encoding neomycin resistance is physically linked to the sequence(s) of the invention. Transfection and isolation of desired clones are carried out by any one of several methods well known to those of ordinary skill in the art (E.J. Robertson, supra).

DNA molecules introduced into ES cells can also be integrated into the chromosome through the process of homologous recombination. Capecchi, Science 244: 1288-1292 (1989). Methods for positive selection of the recombination event (*i.e.*, neo resistance) and dual positive-negative selection (*i.e.*, neo resistance and gancyclovir resistance) and the subsequent identification of the desired clones by PCR have been described by Capecchi, supra and Joyner et al., Nature 338: 153-156, 1989), the teachings of which are incorporated herein. The final phase of the procedure is to inject targeted ES cells into blastocysts and to transfer the blastocysts into pseudopregnant females. The resulting chimeric animals are bred and the offspring are analyzed by Southern blotting to identify individuals that carry the transgene. Procedures for the production of non-rodent mammals and other animals have been discussed by others. See Houdebine and Chourrout, supra; Pursel, et al., Science 244:1281-1288, 1989); and Simms, et al., Bio/Technology 6:179-183, 1988).

Thus, the invention provides transgenic, nonhuman mammals containing a transgene encoding a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide or a gene effecting the expression of a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide. Such transgenic nonhuman mammals are particularly useful as an in vivo test system for studying the effects of introducing a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide, regulating the expression of a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide (*i.e.*, through the introduction of additional genes, antisense nucleic acids, or ribozymes).

A "transgenic animal" is an animal having cells that contain DNA which has been artificially inserted into a cell, which DNA becomes part of the genome of the animal which develops from that cell. Preferred transgenic animals are primates, mice, rats, cows, pigs, horses, goats, sheep, dogs and cats. The transgenic DNA may encode for a human PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide. Native expression in an animal may be reduced by providing an amount of anti-sense RNA or DNA effective to reduce expression of the receptor.

XII. Gene Therapy.

PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 or its genetic sequences, both mutated and non-mutated, will also be useful in gene therapy (reviewed in Miller, Nature 357:455-460, (1992). Miller states that advances have resulted in practical approaches to human gene therapy that have demonstrated positive initial results. The basic science of gene therapy is described in Mulligan, Science 260:926-931, (1993).

In one preferred embodiment, an expression vector containing a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 coding sequence or a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 mutant coding sequence as described above is inserted into cells, the cells are grown in vitro and then infused in large numbers into patients. In another preferred embodiment, a DNA segment containing a promoter of choice (for example a strong promoter) is transferred into cells containing an endogenous PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 in such a manner that the promoter segment enhances expression of the endogenous PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene (for example, the promoter segment is transferred to the cell such that it becomes directly linked to the endogenous PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 gene).

The gene therapy may involve the use of an adenovirus containing PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 cDNA targeted to an appropriate cell type, systemic PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 increase by implantation of engineered cells, injection with PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 virus, or injection of naked PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 DNA into appropriate cells or tissues, for example neurons.

Expression vectors derived from viruses such as retroviruses, vaccinia virus, adenovirus, adeno-associated virus, herpes viruses, several RNA viruses, or bovine papilloma virus, may be used for delivery of nucleotide sequences (*e.g.*, cDNA) encoding recombinant PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 protein into the targeted cell population (*e.g.*., tumor cells or neurons). Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors containing coding sequences. See, for example, the techniques described in Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, N.Y. (1989), and in Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, recombinant nucleic acid molecules encoding protein sequences can be used as naked DNA or in reconstituted system *e.g.*, liposomes or other lipid systems for delivery to target cells (See *e.g.*, Felgner et al., *Nature* 337:387-8, 1989). Several other methods for the direct transfer of plasmid DNA into cells exist for use in human gene therapy and involve targeting the DNA to receptors on cells by complexing the plasmid DNA to proteins. See, Miller, supra.

In its simplest form, gene transfer can be performed by simply injecting minute amounts of DNA into the nucleus of a cell, through a process of microinjection. (Capecchi MR, Cell 22:479-88, 1980). Once recombinant genes are introduced into a

cell, they can be recognized by the cells normal mechanisms for transcription and translation, and a gene product will be expressed. Other methods have also been attempted for introducing DNA into larger numbers of cells. These methods include: transfection, wherein DNA is precipitated with $CaPO_4$ and taken into cells by pinocytosis (Chen C. and Okayama H, Mol. Cell Biol. 7:2745-52, 1987); electroporation, wherein cells are exposed to large voltage pulses to introduce holes into the membrane (Chu G., et al., Nucleic Acids Res., 15:1311-26, 1987); lipofection/liposome fusion, wherein DNA is packaged into lipophilic vesicles which fuse with a target cell (Felgner PL., et al., Proc. Natl. Acad. Sci. USA. 84:7413-7, 1987)); and particle bombardment using DNA bound to small projectiles (Yang NS. et al., Proc. Natl. Acad. Sci. 87:9568-72, 1990). Another method for introducing DNA into cells is to couple the DNA to chemically modified proteins.

It has also been shown that adenovirus proteins are capable of destabilizing endosomes and enhancing the uptake of DNA into cells. The admixture of adenovirus to solutions containing DNA complexes, or the binding of DNA to polylysine covalently attached to adenovirus using protein crosslinking agents substantially improves the uptake and expression of the recombinant gene. Curiel DT et al., Am. J. Respir. Cell. Mol. Biol., 6:247-52, 1992).

As used herein "gene transfer" means the process of introducing a foreign nucleic acid molecule into a cell. Gene transfer is commonly performed to enable the expression of a particular product encoded by the gene. The product may include a protein, polypeptide, anti-sense DNA or RNA, or enzymatically active RNA. Gene transfer can be performed in cultured cells or by direct administration into animals. Generally gene transfer involves the process of nucleic acid contact with a target cell by non-specific or receptor mediated

interactions, uptake of nucleic acid into the cell through the membrane or by endocytosis, and release of nucleic acid into the cytoplasm from the plasma membrane or endosome. Expression may require, in addition, movement of the nucleic acid into the nucleus of the cell and binding to appropriate nuclear factors for transcription.

As used herein "gene therapy" is a form of gene transfer and is included within the definition of gene transfer as used herein and specifically refers to gene transfer to express a therapeutic product from a cell *in vivo* or *in vitro*. Gene transfer can be performed *ex vivo* on cells which are then transplanted into a patient, or can be performed by direct administration of the nucleic acid or nucleic acid-protein complex into the patient.

In another preferred embodiment, a vector having nucleic acid sequences encoding a PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 is provided in which the nucleic acid sequence is expressed only in specific tissue. Methods of achieving tissue-specific gene expression as set forth in International Publication No. WO 93/09236, filed November 3, 1992 and published May 13, 1993.

In all of the preceding vectors set forth above, a further aspect of the invention is that the nucleic acid sequence contained in the vector may include additions, deletions or modifications to some or all of the sequence of the nucleic acid, as defined above.

In another preferred embodiment, a method of gene replacement is set forth. "Gene replacement" as used herein means supplying a nucleic acid sequence which is capable of being expressed in vivo in an animal and thereby providing or augmenting the function of an endogenous gene which is missing or defective in the animal.

XIII. Compounds that Modulate the Function of PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 Proteins.

In an effort to discover novel treatments for diseases, biomedical researchers and chemists have designed, synthesized, and tested molecules that inhibit the function of protein kinases. Some small organic molecules form a class of compounds that modulate the function of protein kinases. Examples of molecules that have been reported to inhibit the function of protein kinases include, but are not limited to, bis monocyclic, bicyclic or heterocyclic aryl compounds (PCT WO 92/20642, published November 26, 1992 by Maguire et al.), vinylene-azaindole derivatives (PCT WO 94/14808, published July 7, 1994 by Ballinari et al.), 1-cyclopropyl-4-pyridyl-quinolones (U.S. Patent No. 5,330,992), styryl compounds (U.S. Patent No. 5,217,999), styryl-substituted pyridyl compounds (U.S. Patent No. 5,302,606), certain quinazoline derivatives (EP Application No. 0 566 266 A1), selenoindoles and selenides (PCT WO 94/03427, published February 17, 1994 by Denny et al.), tricyclic polyhydroxylic compounds (PCT WO 92/21660, published December 10, 1992 by Dow), and benzylphosphonic acid compounds (PCT WO 91/15495, published October 17, 1991 by Dow et al). The compounds that can traverse cell membranes and are resistant to acid hydrolysis are potentially advantageous therapeutics as they can become highly bioavailable after being administered orally to patients. However, many of these protein kinase inhibitors only weakly inhibit the function of protein kinases. In addition, many inhibit a variety of protein kinases and will therefore cause multiple side-effects as therapeutics for diseases.

Some indolinone compounds, however, form classes of acid resistant and membrane permeable organic molecules. PCT WO 96/22976, published August 1, 1996 by Ballinari et al. describes hydrosoluble indolinone compounds that harbor

tetralin, naphthalene, quinoline, and indole substituents fused to the oxindole ring. These bicyclic substituents are in turn substituted with polar moieties including hydroxylated alkyl, phosphate, and ether moieties. International Patent Publication WO 96/22976, published August 1, 1996 by Ballinari et al. describe indolinone chemical libraries of indolinone compounds harboring other bicyclic moieties as well as monocyclic moieties fused to the oxindole ring. WO 96/22976, published August 1, 1996 by Ballinari et al. teach methods of indolinone synthesis, methods of testing the biological activity of indolinone compounds in cells, and inhibition patterns of indolinone derivatives.

Other examples of substances capable of modulating PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity include, but are not limited to, tyrphostins, quinazolines, quinoxolines, and quinolines.

The quinazolines, tyrphostins, quinolines, and quinoxolines referred to above include well known compounds such as those described in the literature. For example, representative publications describing quinazoline include Barker et al., EPO Publication No. 0 520 722 A1; Jones et al., U.S. Patent No. 4,447,608; Kabbe et al., U.S. Patent No. 4,757,072; Kaul and Vougioukas, U.S. Patent No. 5, 316,553; Kreighbaum and Comer, U.S. Patent No. 4,343,940; Pegg and Wardleworth, EPO Publication No. 0 562 734 A1; Barker et al., Proc. of Am. Assoc. for Cancer Research 32:327 (1991); Bertino, J.R., Cancer Research 3:293-304 (1979); Bertino, J.R., Cancer Research 9(2 part 1):293-304 (1979); Curtin et al., Br. J. Cancer 53:361-368 (1986); Fernandes et al., Cancer Research 43:1117-1123 (1983); Ferris et al. J. Org. Chem. 44(2):173-178; Fry et al., Science 265:1093-1095 (1994); Jackman et al., Cancer Research 51:5579-5586 (1981); Jones et al. J. Med. Chem. 29(6):1114-1118; Lee and Skibo, Biochemistry 26(23):7355-7362

(1987); Lemus et al., J. Org. Chem. 54:3511-3518 (1989); Ley and Seng, Synthesis 1975:415-522 (1975); Maxwell et al., Magnetic Resonance in Medicine 17:189-196 (1991); Mini et al., Cancer Research 45:325-330 (1985); Phillips and Castle, J. Heterocyclic Chem. 17(19):1489-1596 (1980); Reece et al., Cancer Research 47(11):2996-2999 (1977); Sculier et al., Cancer Immunol. and Immunother. 23:A65 (1986); Sikora et al., Cancer Letters 23:289-295 (1984); Sikora et al., Analytical Biochem. 172:344-355 (1988).

Quinoxaline is described in Kaul and Vougioukas, U.S. Patent No. 5,316,553.

Quinolines are described in Dolle et al., J. Med. Chem. 37:2627-2629 (1994); MaGuire, J. Med. Chem. 37:2129-2131 (1994); Burke et al., J. Med. Chem. 36:425-432 (1993); and Burke et al. BioOrganic Med. Chem. Letters 2:1771-1774 (1992).

Tyrphostins are described in Allen et al., Clin. Exp. Immunol. 91:141-156 (1993); Anafi et al., Blood 82:12:3524-3529 (1993); Baker et al., J. Cell Sci. 102:543-555 (1992); Bilder et al., Amer. Physiol. Soc. pp. 6363-6143:C721-C730 (1991); Brunton et al., Proceedings of Amer. Assoc. Cancer Rsch. 33:558 (1992); Bryckaert et al., Experimental Cell Research 199:255-261 (1992); Dong et al., J. Leukocyte Biology 53:53-60 (1993); Dong et al., J. Immunol. 151(5):2717-2724 (1993); Gazit et al., J. Med. Chem. 32:2344-2352 (1989); Gazit et al., " J. Med. Chem. 36:3556-3564 (1993); Kaur et al., Anti-Cancer Drugs 5:213-222 (1994); Kaur et al., King et al., Biochem. J. 275:413-418 (1991); Kuo et al., Cancer Letters 74:197-202 (1993); Levitzki, A., The FASEB J. 6:3275-3282 (1992); Lyall et al., J. Biol. Chem. 264:14503-14509 (1989); Peterson et al., The Prostate 22:335-345 (1993); Pillemer et al., Int. J. Cancer 50:80-85 (1992); Posner et al., Molecular Pharmacology 45:673-683 (1993); Rendu et al., Biol. Pharmacology 44(5):881-888 (1992); Sauro and Thomas, Life Sciences 53:371-376 (1993);

Sauro and Thomas, J. Pharm. and Experimental Therapeutics 267(3):119-1125 (1993); Wolbring et al., J. Biol. Chem. 269(36):22470-22472 (1994); and Yoneda et al., Cancer Research 51:4430-4435 (1991).

Other compounds that could be used as modulators include oxindolinones.

Examples

The examples below are non-limiting and are merely representative of various aspects and features of the present invention. The examples below show the isolation and characterization of the novel proteins, protein expression in normal and tumor cells, generation of protein specific antibodies, and recombinant expression in mammalian and yeast systems. Also shown are assays useful for identifying compounds that modulate protein activity.

Example 1: Isolation Of cDNA Clones Encoding PTP04

The example below describes the isolation and identification of a new PTP sequence from primary cancer tissues and the subsequent cloning of a full-length human PTP04. Also described are probes useful for the detection of PTP04 in cells or tissues.

Materials and Methods:

Poly A+ RNA was isolated from 30uM cryostat sections of frozen samples from primary human lung and colon carcinomas (Micro-FastTrack, InVitrogen, San Diego, CA). This RNA was used to generate single-stranded cDNA using the Superscript Preamplification System (GIBCO BRL, Gaithersburg, MD.; Gerard, GF et al. (1989), FOCUS 11, 66) under conditions recommended by the manufacturer. A typical reaction used 10 μg total RNA or 2 μg poly(A) RNA with 1.5 μg oligo$(dT)_{12-18}$ in a reaction volume of

60 μL. The product was treated with RNaseH and diluted to 100 μL with H20. For subsequent PCR amplification, 1-4 μL of this sscDNA was used in each reaction.

Degenerate oligonucleotides were synthesized on an Applied Biosystems 394 DNA synthesizer using established phosphoramidite chemistry, precipitated with ethanol and used unpurified for PCR. The sequence of the degenerate oligonucleotide primers follows:

PTPDFW = 5'-GAYTTYTGGVRNATGRTNTGGGA- (sense) (SEQ ID NO:17) and

PTPHCSA = 5'-CGGCCSAYNCCNGCNSWRCARTG -3' (antisense) (SEQ ID NO:18).

These primers were derived from the peptide sequences DFWXMXW(E/D) (SEQ ID NO:19) (sense strand from PTP catalytic domain) and HCXAGXG (antisense strand from PTP catalytic domain) (SEQ ID NO:20), respectively. Degenerate nucleotide residue designations are: N = A, C, G, or T; R = A or G; and Y = C or T.

PCR reactions were performed using degenerate primers applied to the single-stranded cDNA listed above. The primers were added at a final concentration of 5 μM each to a mixture containing 10 mM Tris·HCl (pH8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM each deoxynucleoside triphosphate, 0.001% gelatin, 1.5 U AmpliTaq DNA Polymerase (Perkin-Elmer/Cetus), and 1-4 μL cDNA. Following 3 min denaturation at 95 °C, the cycling conditions were 94 °C for 30 s, 50 °C for 1 min, and 72 °C for 1 min 45 s for 35 cycles. PCR fragments migrating between 350-400 bp were isolated from 2% agarose gels using the GeneClean Kit (Bio101), and T-A cloned into the pCRII vector (Invitrogen Corp. U.S.A.) according to the manufacturer's protocol.

Colonies were selected for mini plasmid DNA-preparations using Qiagen columns and the plasmid DNA was sequenced using

cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products were run on an ABI Prism 377 DNA Sequencer, and analyzed using the BLAST alignment algorithm (Altschul, S.F. et al., J. Mol. Biol.215:403-10). One novel clone novel clone (G77-4a-117), designated PTP04, was isolated from human HLT370 primary lung carcinoma sample.

To obtain full-length cDNA encoding the novel phosphatase, RACE (rapid amplification of cDNA ends) was performed with sense or anti-sense oligonucleoides derived from the original PCR fragments. Marathon-Ready cDNA (Clontech, Palo Alto, CA) made from human Molt-4 leukemia cells was used in the RACE reactions with the following primers:

RACE primers:

5'-CACCGTTCGAGTATTTCAGATTGTGAAGAAG-TCC-3' (6595) (SEQ ID NO:21),

5'-GGACTTCTTCACAATCTGAAATACTCGAACGGTG-3' (6596) (SEQ ID NO:22),

5'-CCGTTATGTGAGGAAGAGCCACATTACAGGACC-3' (6599) (SEQ ID NO:23),

5'-GGTCCTGTAATGTGGCTCTTCCTCACATAACGG-3' (6600) (SEQ ID NO:24),

AP-1, and AP-2 (Clontech).

RT-PCR primers for PTP04:

5'-GGCATGCATGGAGTATGAAATGG-3' (6618) (SEQ ID NO:25),

5'-CGTACATCCCAGATGAGCTCAAGAATAGGG-3' (6632) (SEQ ID NO:26).

Isolated cDNA fragments encoding PTP04 were confirmed by DNA sequencing and subsequently used as probes for the screening of a human leukocyte cDNA library.

A human leukocyte cDNA library (lTriplEx, Clontech) and a Molt-4 leukemia cell library (lgt11, Clontech) were then screened to isolate full-length transcripts encoding PTP04. The 5' or 3'-RACE fragments were ^{32}P-labeled by random priming and used as hybridization probes at $2x10^6$ cpm/mL following standard techniques for library screening. Pre-hybridization (3 h) and hybridization (overnight) were conducted at 42 °C in 5X SSC, 5 X Denhart's solution, 2.5% dextran sulfate, 50 mM $Na_2PO_4/NaHPO_4$ [pH 7.0], 50% formamide with 100 mg/mL denatured salmon sperm DNA. Stringent washes were performed at 65 °C in 0.1X SSC and 0.1% SDS. Several overlapping clones were isolated and found to span the sequence of the PCR fragment (G77-4a-117). The final sequence was verified by sequencing of both strains using a cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products were run on an ABI Prism 377 DNA Sequencer.

Results:

The 3,580 bp human PTP04 nucleotide sequence encodes a polypeptide of 807 amino acids. The PTP04 coding sequence is flanked by a 52 nucleotide 5'-untranslated region and a 1086 nucleotide 3'-untranslated region ending with a poly(A) tail. While there are no upstream in frame stop codons, the first ATG beginning at nucleotide position 53 conforms to the Kozak consensus for an initiating methionine. This predicted first 6 amino acids are identical to those of murine ZPEP (SwissProt: P29352, GeneBank: M90388), further supporting this is the true translational start site. One cDNA clone had an insert after nucleotide 30 in the 5'UTR, but otherwise had no sequence differences.

The 807 amino acid sequence shows no signal sequence or a transmembrane domain and PTP04 is, therefore, an intracellular protein. PTP04 has an N-terminal region from amino acids 1-48,

a catalytic domain from amino acids 49-294, and a C-terminal tail from amino acids 295-807. PTP04 is most related to murine ZPEP with an overall homology of 70%. ZPEP is a member of a subfamily of PTPs that includes PTP-PEST, HSC, BDP1 and PTP20, all of which are cytoplasmic PTPs with a single catalytic domain and a region rich in Pro, Glu, Ser and Thr residues (PEST domain). PTP04 also has a C-terminal PEST domain, from amino acids 495-807, where there are 57 serine residues (18%) and 35 proline residues(11%). A comparison of the amino acid sequences of PTP04 and ZPEP is shown in Figure 1.

The homology between PTP04 and ZPEP is concentrated in the N terminal and C-terminal ends of the proteins with significant divergence in the middle. The N-terminal region of PTP04, from amino acids 1-48, is 81% homologous to murine ZPEP. The catalytic domain of PTP04, from amino acids 49-294, is 89% homologous to murine ZPEP. The region of PTP04 from 294-600 is approximately 50% homologous to murine ZPEP. The C-terminal region of PTP04, from 680-817, is 80% homologous to murine ZPEP. The human SuPTP04 sequence defines a novel member of the PTP-PEST subfamily of PTPs.

Example 2: Expression Of PTP04

The example below shows the evaluation of PTP04 expression in normal human tissues and in cancer cell lines.

Materials and Methods:

Northern blots were prepared by running 20 μg total RNA per lane isolated from 22 human adult normal tissues (thymus, lung, duodenum, colon, testis, brain, cerebellum, salivary gland, heart, liver, pancreas, kidney, spleen, stomach, uterus, prostate, skeletal muscle, placenta, mammary gland, bladder, lymph node, adipose tissue), 2 human fetal normal tissues (fetal liver, fetal brain), and 24 human tumor cell lines (

HOP-92, EKVX, NCI-H23, NCI-H226, NCI-H322M, NCI-H460, A549, HOP-62, OVCAR-3, OVCAR-4, OVCAR-5, OVCAR-8, IGROV1, SK-OV-3, SNB-19, SNB-75, U251, SF-268, SF-295, SF-539, CCRF-CEM, SR, DU-145, PC-3) (obtained from Nick Scuidero, National Cancer Institute, Developmental Therapeutics Program, Rockville, MD). The total RNA samples were run on a denaturing formaldehyde 1% agarose gel and transferred onto a nitrocellulose membrane (BioRad, CA). An additional human normal tissue Northern blot containing 2 μg polyA+ mRNA per lane from 8 different human cancer cell lines (NCI-H522, K-562, MOLT-4, HL-60, S3, Raji, SW480, G361) on a charge-modified nylon membrane (human cancer cell line blot #7757-1, Clontech, Palo Alto, CA) were also hybridized.

For the total RNA samples, nitrocellulose membranes were hybridized with randomly primed [a-^{32}P]dCTP-labeled probes synthesized from a 579 bp StuI-BstXI fragment of pCR2.1.mini298. Hybridization was performed overnight at 42°C in 4X SSPE, 2.5X Denhardt's solution, 50% formamide, 0.2 mg/mL denatured salmon sperm DNA, 0.1 mg/mL yeast tRNA (Boehringer Mannheim,IN), 0.2% SDS, with 5 x 10^6 cpm/mL of [a-^{32}P]dCTP labeled DNA probes on a Techne hybridizer HB-1. The blots were washed with 2X SSC, 0.1% SDS, at 65 °C for 20 min twice followed by in 0.5 X SSC, 0.1% SDS at 65 °C for 20 min. The blots were exposed to a phospho-imaging screen for 24 hours and scanned on a Molecular Dynamics Phosphoimager SF.

A 351 bp EcoRI-HindIII fragment of G77-4a-117 was used to generate a probe for 2 μg poly A+ mRNA samples on a Clontech nylon membrane. Hybridization was performed at 42 °C overnight in 5X SSC, 2% SDS, 10X Denhardt's solution, 50% formamide, 100 μg/mL denatured salmon sperm DNA with 1-2 x 10^6 cpm/mL of [a-^{32}P]dCTP -labeled DNA probes. The membrane was washed at room temperature in 2X SSC/0.05% SDS for 30 min and followed by at

50 °C in 0.2X SSC/0.1% SDS for 30 min, twice, and exposed for 48 hours on Kodak XAR-2 film.

RT-PCR Detection of novel PTPs -

Total RNA was isolated from various cell lines or fresh frozen tissues by centrifugation thrugh a cesium chloride cushion. Twenty μg of total RNA was reverse transcribed with random hexamers and Moloney murine leukemia virus reverse transcriptase (Super-ScriptII, GIBCO BRL, Gaithersburg, MD). PCR was then used to amplify cDNA encoding SuPTP04. RT-PCR reactions lacking only the reverse transcriptase were performed as controls. PCR products were electrophoresed on 3% agarose gels, visualized by ethidium bromide staining and photographed on a UV light box. The intensity for a 270-bp fragment specific to PTP04 were compared among different RNA samples.

Results:

A single SuPTP04 mRNA transcript of approximately 4.5 kb was identified by Northern analysis, and found to be exclusively in the Thymus. The rest of 23 human normal tissues (fetal brain, fetal liver, lung, duodenum, colon, testis, brain, cerebellum, salivary gland, heart, liver, pancreas, kidney, spleen, stomach, uterus, prostate, skeletal muscle, placenta, mammary gland, bladder, lymph node, adipose tissue) were all negative. Six of the human tumor cell lines (CCRF-CEM, K-562, MOLT-4, HL-60, SR, Raji) were positive. The rest of 26 human tumor cell lines were negative. RT-PCR with gene specific primer-pairs showed that expression of the transcripts encoding SuPTP04 confirmed the results from Northern analysis and also detected low levels in adipose, kidney, small intestine, hematopoietic tissues and various cell types (spleen, thymus, lymph node, bone marrow, peripheral leukocytes and lymphocytes.

The selective expression of PTP04 in cells of hematopoetic origin including normal human thymus and several leukemia cell lines suggests a potential involvement in immune regulation including T and B cell survival, differentiation or co-stimulation, and/or inflammatory, immunosuppressive or autoimmune disorders. Additionally, expression in adipose tissue suggests a possible role in metabolic disorders such as diabetes.

Example 3: Recombinant Expression Of PTP04

The following example illustrates the construction of vectors for expression of recombinant PTP04 and the creation of recombinant cell lines expressing PTP04.

Construction of Expression Vectors -

Expression constructs were generated by PCR-assisted mutagenesis in which the entire coding domains of PTP04 was tagged on its carboxy-terminal end with the hemophilus influenza hemaglutinin (HA) epitope YPYDVPDYAS (SEQ ID NO:55) (Pati, 1992). The construct was introduced into two mammalian expression vectors: pLXSN (Miller, A.D. & Rosman, G.J., Biotechniques 7, 980-988, 1989) for the generation of virus producing lines; and pRK5 for transient expression in mammalian.

Dominant negative (signaling incompetent) PTP04 constructs were also made in both pLXSN and pRK5 by mutation of the invariant Cys in the conserved HCSAG (SEQ ID NO:56) motif to an Ala by PCR mutagenesis.

The entire PTP04 open reading frames (no HA-tag) excluding the initiating methionines were generated by PCR and ligated into pGEX vector (Pharmacia Biotech, Uppsala, Sweden) for bacterial production of GST-fusion proteins for immunization of rabbits for antibody production. The entire PTP04 open reading

frame excluding the initiating methionines was generated by PCR and ligated into pGEX vector for bacterial production of GST-fusion proteins for immunization of rabbits for antibody production. This vector contains the glutathione-S-transferase coding sequence followed by a polylinker for generating recombinant fusion proteins. The GST moiety comprises the N-terminal portion of the fusion protein.

Transient Expression in Mammalian Cells -

The pRK5 expression plasmids (10 μg DNA/100 mm plate) containing the HA-tagged PTP04 gene can be introduced into COS and 293 cells with lipofectamine (Gibco BRL). After 72 hours, the cells were harvested in 0.5 mL solubilization buffer (20 mM HEPES pH 7.35, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1.5 mM $MgCl_2$, 1 mM EGTA, 2 mM phenylmethylsulfonyl fluoride, 1 μg/mL aprotinin). Sample aliquots were resolved by SDS polyacrylamide gel electrophoresis (PAGE) on 15%acrylamide/0.5% bis-acrylamide gels and electrophoretically transferred to nitrocellulose. Non-specific binding was blocked by preincubating blots in Blotto (phosphate buffered saline containing 5% w/v non-fat dried milk and 0.2% v/v nonidet P-40 (Sigma)), and recombinant protein was detected using a murine Mab to the HA decapeptide tag. Alternatively, recombinant protein can be detected using various PTP04-specific antisera.

Generation of Virus Producing Cell Lines

pLXSN recombinant constructs containing the PTP04 gene were transfected into an amphotropic helper cell line PA317 using $CaCl_2$ mediated transfection. After selection on G418, the cells were plated on normal media without G418 (500 μg/mL). Supernatants from resistant cells were used to infect the ecotropic helper cell line GP+E86, and cells again selected on

G418. Resistant cells were again taken off G418, and the supernatants harvested every 8-12 hours and pooled as virus stock. Redemann et al., 1992, Mol. Cell. Biol. 12: 491-498. Viral stock titers were typically ~10^6/mL.

Stable Expression in Mammalian Cells

NIH-3T3, and BALB/3T3 cells were grown in 100 mm plates with DMEM (Gibco) containing 10% fetal calf serum (FCS). The cells were superinfected with the PTP04 retrovirus by adding approximately 3 mL viral supernatant to 15 mL culture media for approximately 24 hours. Cells expressing the retroviral constructs were then selected by growth in DMEM/10% FCS supplemented with 500 μg/mL G418.

Example 4: Generation of Anti-PTP04 Antibodies

PTP04-specific immunoreagents were raised in rabbits against a mixture of three KLH-conjugated synthetic peptides corresponding to unique sequences present in human PTP04. The peptides (see below) were conjugated at the C-terminal residue with KLH.

peptide 428A: SWPPSGTSSKMSLDDLPEKQDGTVFPSSLLP (SEQ ID NO:27)

peptide 429A: YSLPYDSKHQIRNASNVKHHDSSALGVYSY (SEQ ID NO:28)

peptide 430A: HTLQADSYSPNLPKSTTKAAKMMNQQRTKC (SEQ ID NO:29)

Additional immunoreagents were generated by immunizing rabbits with the bacterially expressed entire coding region of PTP04 expressed as a GST-fusion protein. GST fusion proteins were produced in DH5-alpha E. coli bacteria as descaribed in

Smith, et al Gene 67:31, 1988. Bacterial protein lysates were purified on glutathione-sepharose matrix as described in Smith, et al, supra.

Example 5: Assay for PTP04 Activity

Materials and methods:

Recombinant wild-type and dominant negative (signaling incompetent) PTP04 (see Example 3, supra) were purified from bacteria as GST-fusion proteins. Lysates were bound to a glutathione-sepharaose matrix and washed twice with 1X HNTG, followed by one wash with a buffer containing 100 mM 2-(N-morpholino)ethansulfonic acid (MES), pH 6.8, 150 mM NaCl, and 1 mM EDTA.

The assay for phosphatase activity was essentially done as described by Pei et al.(1993) using p-nitrophenolphosphate (PNPP) as a generic PTP substrate. Briefly, after the last washing step, reactions were started by adding 50 mL Assay Buffer (100 mM MES pH 6.8, 150 mM NaCl, 10 mM DTT, 2 mM EDTA, and 50 mM PNPP) to the matrix bound proteins. Samples were incubated for 20 min. at 23 °C. The reactions were terminated by mixing 40 μL of each sample with 960 μL 1 N NaOH, and the absorbance of p-nitrophenol was determined at 450 nm. To control for the presence of PTP04 in the precipitates, the precipitates were boiled in SDS sample buffer and analyzed by SDS-PAGE. The presence of PTP04 was then detected by immunoblot analysis with anti-PTP04 antibodies.

Example 6: Isolation and Characterization of SAD

This example describes the isolation and characterization of the non-receptor tyrosine kinase SAD. Initially we set out to identify novel members of the Src family, a group of nine related cytoplasmic tyrosine kinases which play key roles in several signal transduction pathways. Based on comparison of

all known tyrosine kinases, we designed a pair of degenerate oligonucleotide primers that specifically recognize Src family members plus three more distantly related proteins Srm, Brk, and MKK3 or Frk (the Srm/Brk/Frk group). The sequence FGE/DVW (SEQ ID NO:30) is located near the amino terminus of the kinase domain and is unique to Src family members and the Srm/Brk/Frk group. The sequence WTAPE (SEQ ID NO: 31) is located just amino terminal to the highly conserved DVWS motif of tyrosine kinases and is contained in the Src family and the Srm/Brk/Frk group as well as the Eph, Csk, Abl, and Fes families.

When we used the FGE/DVW and WTAPE primers in PCR amplification reactions with HME (human mammary epithelial) cell sscDNA as a template, we isolated multiple copies of known Src relatives as well as a novel DNA fragment (HME 1264) of 483 bp with homology to other kinases. The novel sequence was most similar to mouse Srm (GeneBank Accession #D26186) and the clone was designated human SAD.

A SAD probe was used to screen a cDNA library constructed from human breast cancer cell line mRNA to isolate two overlapping, independent clones spanning the kinase domain, but containing apparent introns and presumably arising from incompletely processed transcripts. The 5' end of the coding region was subsequently isolated by sequential RACE reactions from MCF7 RNA, and the entire coding region was re-isolated by PCR from HME cDNA.

Materials And Methods

Total RNA was isolated using the Guanidine Salts/Phenol extraction protocol of Chomczynski and Sacchi (P. Chomczynski and N. Sacchi, Anal. Biochem. 162, 156 (1987) from HME (human mammary epithelial) cells. This RNA was used as a template to generate single-stranded cDNAs using the Superscript Pre-amplification System for First Strand Synthesis kit purchased

from GibcoBRL (Life Technologies, U.S.A.; Gerard, GF et al, FOCUS 11:66, 1989) under conditions recommended by manufacturer. A typical reaction used 10 μg total RNA or 2 μg poly(A)$^+$ RNA with 1.5 μg oligo(dT)$_{12-18}$ in a reaction volume of 60 μL. The product was treated with RNaseH and diluted to 100 uL with H_2O. For subsequent PCR amplification, 1-4 μL of these sscDNAs were used in each reaction.

Oligonucleotides were synthesized on an Applied Biosystems 394 DNA synthesizer using established phosphoramidite chemistry and were used unpurified after precipitation with ethanol. The degenerate oligonucleotide primers are:

FGE/DVW = 5'-GGNCARTTYGGNGANGTNTGG-3' (SEQ ID NO:30) (sense)

and

WTAPE = 5'-CAGNGCNGCYTCNGGNGCNGTCCA-3' (SEQ ID NO:31) (antisense).

These primers were derived from the peptide sequences GQFG(E/D)VW (SEQ ID NO:32) (sense strand) and WTAPEALL (SEQ ID NO:33) (antisense strand), respectively. Degenerate nucleotide residue designations are: N = A, C, G, or T; R = A or G; and Y = C or T. Using Src as a template, these primers produce a product of 480 bp.

A PCR reaction was performed using primers FGE/DVW and WTAPE applied to HME cell cDNA. The primers were added at a final concentration of 0.5 uM each to a mixture containing 10 mM Tris.HCl (pH8.3), 50 mM KCl, 1.5 mM MgCl2, 200 uM each deoxynucleoside triphosphate, 0.001% gelatin, and 1.5 U AmpliTaq DNA Polymerase (Perkin-Elmer/Cetus), and 1-4 µL cDNA. Following 3 min denaturation at 94 °C, the cycling conditions were 94 °C for 30 sec, 37 °C for 1 min, a 2 min ramp to 72 °C, and 72°C for 1min for the first 3 cycles, followed by 94 °C for 30 sec, 60°C for 1 min, and 72 °C for 1 min for 35 cycles. PCR fragments migrating at between 450-550 bp were isolated from 2%

agarose gels, phosphorylated and repaired by treatment with T4 polynucleotide kinase and Klenow fragment, and blunt-end cloned into the EcoRV site of the vector pBluescriptSK+ (Stratagene U.S.A.).

Plasmid DNAs were isolated from single colonies by DNA minipreparations using QIAGEN columns and were sequenced using cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products were run on an ABI Prism 377 DNA Sequencer, and analyzed using the BLAST alignment algorithm (Altschul, S.F. et al., J. Mol. Biol.215:403-10, 1990). A novel clone (HME1264) was isolated by PCR with primers FGE/DVW and WTAPE on single-stranded cDNA from HME cells as a template. This clone was subsequently designated as a fragment of human SAD.

A lambda ZapII (Stratagene Cloning Systems, La Jolla, CA) cDNA library was constructed using mRNA from a pool of breast carcinoma cell lines as a template for first strand cDNA synthesis with both oligo-(dT) and random priming (library created by Clonetech custom library synthesis department, Palo Alto, CA). The cell lines used for the pool were MCF7, HBL100, MDA-MB231, MDA-MB175IIV, MDA-MB435, MDA-MB453, MDA-MB468, BT20, T47D and SKBR3, all of which are available from the ATCC. Phage were screened on nitrocellulose filters with the random primed ^{32}P-labeled insert from HME1264 at $2x10^6$ cpm/mL in hybridization buffer containing 6xSSPE, 50% formamide, 2x Denhardt's reagent, 0.1% SDS, with 0.05 mg/mL denatured, fragmented salmon sperm DNA. After overnight hybridization at 42 °C, filters were washed in 1xSSC, 0.1% SDS at 65 °C. Two overlapping partial clones were isolated and sequenced through the coding region using manual sequencing with T7 polymerase and oligonucleotide primers (Tabor and Richardson, Proc. Natl. Acad. Sci. U.S.A. 84: 4767-71, 1987). These isolates encompass the kinase domain of SAD and extend from within an apparent

intron 5' to the kinase domain and extend 3' to an in-frame termination codon, but are interrupted by four more apparent introns.

Two sequential 5' RACE (rapid amplification of cDNA ends) reactions (Frohman et al., Proc. Natl. Acad. Sci. U.S.A. 85: 8998, 1988) were subsequently used to isolate the 5' end of the coding region. Single strand cDNA was prepared as described above using the Superscript Pre-amplification System (GibcoBRL) using 6 μg total RNA from MCF7 cells as a template and gene specific primers 5556 (5'-AGTGAGCTTCATGTTGGCT-3' (SEQ ID NO:34) for RACE 1 or 5848 (5'-GGTAGAGGCTGCCATCAG-3' (SEQ ID NO:35)) for RACE 2 to prime reverse transcription. Following treatment with RNase H, sscDNA was recovered using two sequential ethanol precipitations with ammonium acetate and carrier glycogen homopolymer tail of dA was added by treatment with deoxy-terminal transferase (GibcoBRL) and two reaction mixtures diluted to 50 μL with TE. Second strand cDNA synthesis by AmpliTaq DNA Polymerase (Perkin-Elmer/Cetus) was primed with 0.2 uM PENN$(dT)_{17}$ (5'- GACGATCGGAATTCGCGA$(dT)_{17}$-3' (SEQ ID NO:36) using 1-5 μL of tailed cDNA as a template and buffer conditions given above. Following 5 min denaturation at 94 °C, 1 min annealing at 50 °C, and 40 min extension at 72 °C, primers PENN (5'-GACGATCGGAATTCGCGA-3' (SEQ ID NO:37) and 5555 (5'-CCCAGCCACAGGCCTTC-3' (SEQ ID NO:38) were added at 1 µM and PCR done with cycling conditions of 94 °C for 30 s, 49 °C for 1 min, and 72 °C for 1 min, 45 sec for 40 cycles. A second, nested PCR was done using 0.2 µL of the initial PCR reaction as a template and primers PENN (see SEQ ID NO:37) and 5554 (5'-CCACACCTCCCCAAAGTA-3' (SEQ ID NO:39) at 1 μM with an initial 3 min denaturation at 94 °C, followed by cycling conditions of 94 °C for 30 s, 49 °C for 1 min, and 72 °C for 1 min, 45 sec for 35

cycles. PCR products were separated on 1% agarose gels and visualized by ethidium bromide staining and Southern hybridization using oligonucleotide 5557 (5'-TGGGAGCGGCCACACTCCGAATTCGCCCTT-3' (SEQ ID NO:40) end-labeled with ^{32}P. Reaction products of 500-700 bp were digested with EcoRI and cloned into the EcoRI site of pBluescriptSK+ (Stratagene U.S.A.), and positive clones were identified by colony hybridization with oligonucleotide 5557 as a probe. Clone 16A1 (which encompasses nucleotides 195 to 783 of SEQ ID NO:10) was isolated and sequenced by a combination of ABI and manual sequencing.

A second set of 5' RACE reactions was done based on the sequence of clone 16A1 using the procedure given above except as noted. Gene specific primers were 5848 (SEQ ID NO:35) for the first strand synthesis, 6118 (5'-GCCTGCGTGCGAAGATG-3' (SEQ ID NO:41) for the first PCR, and 6119 (5'-CTTCGAGGGCACAGAGCC-3' (SEQ ID NO:42) for the second PCR, and the probe for Southern and colony hybridization was random primed 32P-labeled insert from 16A1. PCR fragments migrating at between 250-450 bp were isolated from 2% agarose gels, phosphorylated and repaired by treatment with T4 polynucleotide kinase and Klenow fragment, and blunt-end cloned into the EcoRV site of the vector pBluescriptSK+ (Stratagene U.S.A.). Clone 20E2 (which encompasses nucleotides 1 to 267 of SEQ ID N0:10) was isolated and sequenced by a combination of ABI and manual sequencing.

The coding region of SAD was subsequently isolated from HME cDNA as two overlapping PCR fragments. Single stranded cDNA was prepared from poly(A)+ RNA from HME cells using the Superscript Preamplification System (GibcoBRL) as described above. PCR with AmpliTaq DNA Polymerase (Perkin-Elmer/Cetus) used 1-2 μL of cDNA as a template, an initial 3 min denaturation at 94°C, followed by cycling conditions of 94oC for

30 s, 55 °C for 1 min, and 72 °C for 1 min, 45 sec for 30 cycles and the buffer conditions given above. Primers 6642 (5'-ATGGAGCCGTTCCTCAGGAGG-3' (SEQ ID NO:43) and 6644 (5'-TCACCCAGCTTCCTCCCAAGG-3' (SEQ ID NO:44) were used to amplify an approximately 710 bp 5' fragment of SAD, and primers 6643 (5'-AGGCCAACTGGAAGCTGATCC-3' (SEQ ID NO:45) and 6645 (5'-GCTGGAGCCCAGAGCGTTGG-3' (SEQ ID NO:46) were used to amplify an approximately 860 bp 3' fragment of SAD. PCR fragments were isolated from 1% agarose gels, phosphorylated and repaired by treatment with T4 polynucleotide kinase and Klenow fragment, and blunt-end cloned into the EcoRV site of the vector pBluescriptSK+ (Stratagene U.S.A.). Positive clones were identified by colony hybridization with the random primed 32P-labeled insert from 16A1 (for the 5' fragment) and the random primed 32P-labeled insert from HME1264 or 32P-labeled oligonucleotide 5557 (for the 3' fragment) as probes. The overlapping 5' and 3' PCR fragments were ligated together via the unique EcoRI site to give the full length SAD coding region. The complete sequence of the coding region of huma SAD was determined from overlapping 5' and 3' PCR clones amplified from cDNA prepared from HME cells. 5' noncoding sequence was determined from the overlapping RACE fragment 16A1. Sequence was determined manually on both strands using cycle sequencey dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA).

Results

The 1,548 bp human SAD (SAD_h) nucleotide sequence shown in SEQ ID NO:10 contains a single open reading frame encoding a polypeptide of 488 amino acids. The SAD_h coding region is preceded by a 48 nucleotide 5'-untranslated region including an in-frame termination codon four codons before the initiating

methionine and a 33 nucleotide 3'-untranslated region that includes two in-frame stop codons.

The sequences of SAD cDNAs were determined from overlapping PCR-amplified fragments from normal HME cell cDNA (nucleotides 49-1548), clones from a breast carcinoma cell lambda cDNA library (nucleotides 694-1548), and overlapping 5' RACE products from MCF7 cDNA (nucleotides 1-783) with the following sequence differences including some likely polymorphic sites. Ambiguities include a change of nucleotide 636 (see SEQ ID NO:10) from a C in the HME PCR clone to a T in the MCF& RACE product, nucleotide 1477 from a T in the HME PCR clone to a C in the breast carcinoma libray, a deletion of nucleotides GT at positions 919 - 920 in the breast carcinoma library and apparent introns inserted at positions (relataive to SEQ ID NO:10) 694, 995, 1117, and 1334 in the breast carcinoma library.

The domain structure of SAD consists of an N-terminal unique domain followed by an SH3 domain, an SH2 domain and a kinase domain. This overall topology is shared by members of the Src, Srm/Brk/Mkk3, and Csk families. SAD is most similar to mouse Srm (GeneBank Accession #D26186) (Kohmura et al., Mol. Cell. Biol. 14: 6915-6925, 1994), a distant SRC relative of unknown function. SAD and Srm share sequence identities in the individual domains of 55% (unique region), 72% (SH3 domain), 78% (SH2 domain), and 85% (kinase domain). Unlike true Src family members, SAD and Srm lack both an N-terminal membrane attachment sequence and a potential C-terminal negative regulatory tyrosine. In addition, the characteristic HRDLRXAN (SEQ ID NO:47) sequence in the Src family kinase domain is HRDLAXRN (SEQ ID NO:48) in SAD and other Srm/Brk/Mkk3 group members. Like most other tyrosine kinases, SAD and Srm both contain a potential autophosphorylation site (380Y of SAD). The N-terminal sequences of SAD and Srm are similar with twenty

identical residues out of the first twenty-two amino acids, but the extreme C-termini are quite distinct.

Available evidence suggests that SAD_h and Srm_m are distinct genes rather than species orthologues. Overall, the levels of homology between SAD_h and Srm_m listed above are comparable to those of close Src family members (for example Src_h and Yes_h), but lower than those of species counterparts (for example Src_h and Src_m). SAD_h and Srm_m also exhibit distinct expression patterns with SAD_h expression detected by PCR in the duodenum and perhaps testes, but not in other tissues tested, while the Srm_m expression was detected by Northern with highest levels in lung, liver, spleen, kidney, and testes (Kohmura et al., Mol. Cell. Biol. 14: 6915-6925, 1994) (See Example 2 below.). Lastly, disruption of the Srm gene in mice has no detectable phenotype (Kohmura et al., Mol. Cell. Biol. 14: 6915), suggesting that other related proteins might compensate for its function.

Example 7: SAD Expression Analysis

Materials And Methods

RNA was isolated from a variety of human cell lines and fresh frozen normal tissues. (Tumor cell lines were obtained from Nick Scuidero, National Cancer Institute, Developmental Therapeutics Program, Rockville, MD)Single stranded cDNA was synthesized from 10 µg of each RNA as described above using the Superscript Preamplification System (GibcoBRL). These single strand templates were then used in a 35 cycle PCR reaction using an annealing temperature of 65 °C with two SAD-specific oligonucleotides (5284: 5'-TCGCCAAGGAGATCCAGACAC-3' (SEQ ID NO:49), and 5285: 5'-GAAGTCAGCCACCTTGCAGGC-3' (SEQ ID NO:50). Reaction products were electrophoresed on 2% agarose gels,

stained with ethidium bromide and photographed on a UV light box. The relative intensity of the approximately 320-bp SAD-specific band was estimated for each sample. The results are shown with a numerical rating with 4 being the highest relative expression and 0 being the lowest.

Results

The SAD expression profile in normal human tissue and multiple cell lines of diverse neoplastic origin was determined by the semi-quantitative PCR assay using primers from sequences in the kinase domain. The results are included in Tables 1 and 2. In normal tissue samples (Table 1), modest SAD expression was detected in the duodenum and possible low levels in testes with all other samples negative. Much higher expression was found in a subset of cancer cell lines (Table 2) with the highest levels in some human colon tumor cell lines (HCT-15, SW480, and HT-29), an ovarian carcinoma (IGROV1), and an intestinal carcinoma (SNU-C2B). Lesser expression of SAD was also seen in some other tumor cell lines derived from colon, breast, lung, ovary, and kidney as shown in Table 2.

Table 1

cell type	Origin	exp. level
duodenum	Normal Tissue	2
testes	Normal Tissue	1?
brain	Normal Tissue	0
heart	Normal Tissue	0
kidney	Normal Tissue	0
lung	Normal Tissue	0
pancreas	Normal Tissue	0
placenta	Normal Tissue	0
salivary gland	Normal Tissue	0
skeletal muscle	Normal Tissue	0
spleen	Normal Tissue	0
stomach	Normal Tissue	0
thymus	Normal Tissue	0
cerebellum	Normal Tissue	0
liver	Normal Tissue	0
uterus	Normal Tissue	0
prostate	Normal Tissue	0

Table 2

Cell Line	Origin	exp.	Cell Line	Origin	exp.
HCT-15	colon	4	LOX IMVI	melanoma	1?
IGR0V1	ovary	4	KATO III	gastric carcinoma	0
SW480	colon adenocarcinoma	3	R-48	meta gast. adenocarcinoma	0
SNU-C2B	cecum primary carcinoma	3	HFL1	lung, diploid	0
HT-29	colon	3	HOP62	lung	0
Colo 205	colon carcinoma	2	OVCAR-4	ovary	0
SW948	colon adenocarcinoma	2	SKOV3	ovary	0
HCT116	colon	2	NCIH23	lung	0
EKVX	lung	2	NCI-H460	lung	0
NCI-H23	lung	2	COLO205	colon	0
HCC-2998	colon	2	NCI-H460	lung	0
HCT116	colon	2	A549/ATCC	LUNG	0
MCF7	breast	2	HOP-62	lung	0
T-47D	breast	2	COLO 205	colon	0
OVCAR-3	ovary	2	KM-12	colon	0
OVCAR-5	ovary	2	MDA-MB-231	breast	0
OVCAR-8	ovary	2	MDA-MB-435	breast	0
SN12C	renal	2	MDA-N	breast	0
ACHN	renal	2	BT-549	breast	0
786-0	renal	2	SNB-19	CNS	0
TK-10	renal	2	SNB-75	CNS	0
HT29	colon adenocarcinoma	1	U251	CNS	0
RF-1	gastric carcinoma	1	SF-268	CNS	0
AGS	gastric carinoma	1	SF-295	CNS	0

EKVX	lung	1	CCRF-CEM	leukemia	0
HOP-92	lung	1	MOLT-4	leukemia	0
NCI-H226	lung	1	HL-60(TB)	leukemia	0
NCI-H322M	lung	1	RPMI8226	leukemia	0
MCF7/ADR	breast	1	SR	leukemia	0
OVCAR-4	ovary	1	UO-31	renal	0
SF-539	CNS	1	A498	renal	0
K-562	leukemia	1	Caki-1	renal	0
RXF393	renal	1	SK-MEL-2	melanoma	0
Calu-3	lung adenocarcinoma	1?	SK-MEL-5	melanoma	0
NCI-H522	lung	1?	SK-MEL-28	melanoma	0
SW620	colon	1?	UACC-62	melanoma	0
Hs578T	breast	1?	UACC-257	malanoma	0
Sk-OV-3	ovary	1?	M14	melanoma	0

Example 8: Generation of SAD-specific Immunoreagents

A SAD-specific antisera was raised in rabbits against a KLH-conjugated synthetic peptide derived from the C-terminal region of SAD (amino acids 478 to 488 of SEQ ID NO:35) with a C to S substitution at position 486 essentially as described in Gentry and Lawton, Virology 152:421, 1984.

Example 9: Recombinant Expression of SAD

Construction Of Vectors

Expression constructs were generated by PCR-based mutagenesis in which a BamHI site was introduced upstream of the initiating Met giving a 5' untranslated sequence of 5'-GGATCCCCGGACC-3' (SEQ ID NO:51). An N-terminal hexahistidine tagged construct was also created by PCR with the coding sequence for MRGSHHHHHH (SEQ ID NO:52) (ATGAGAGGATCGCATCACCATCACCATCAC, SEQ ID NO: 53) followed by the second nucleotide of the SAD coding sequence (a glutamate). Proteins tagged with this sequence can be recognized by the RGS●His Antibody (QIAGEN Inc.) and affinity purified with Ni-NTA resin (QIAGEN Inc.). These constructs were cloned into the 5'-BamHI and 3'-EcoRI sites of pBluescriptSK+ (Stratagene U.S.A.) and the 5'-BamHI and 3'-XhoI sites of the mammalian expression pcDNA3 (Invitrogen) for transient expression analysis.

The SpeI-XhoI full length SAD constructs were also cloned from pBluescriptSK+ (Stratagene U.S.A.) into the yeast expression vector pRSP (Superti-Furga et al., EMBO J. 12, 2625-2634). This vector contains a thiamine-repressible promoter in a shuttle vector for inducible expression in Schizo-saccharomyces pombe. This system has been useful in studies of SRC family members for testing negative regulation by CSK, screening for additional regulators, and purifying recombinant

protein (Superti-Furga et al., EMBO J. 12, 2625-2634; Superti-Furga et al., Nature Biotech. 14, 600-605).

Transient Expression of SAD in Mammalian Cells

The pcDNA3 expression plasmids (5 µg DNA/60 mm plate) containing the unmodified and hexahistidine-tagged SAD genes were introduced into 293 cells with lipofectamine (Gibco BRL). After 48 hours, the cells were harvested in 0.25 mL RIPA (20 mM Tris-Cl pH 7.5, 150 mM NaCl, 1% Triton X-100, 1% deoxycholate, 0.1% SDS, 1mM DTT, 1 mM sodium vanadate, 1 mM phenylmethylsulfonyl fluoride, 2 µg/mL aprotinin, 1 µg/mL leupeptin, and 25 µg/mL trypsin inhibitor). Sample aliquots were resolved by SDS polyacrylamide gel electrophoresis (PAGE) on 10% acrylamide gels and electrophoretically transferred to nitrocellulose. Non-specific binding was blocked by preincubating blots in Blotto (Tris buffered saline containing 5% w/v non-fat dried milk and 0.1% v/v Tween-20), and recombinant protein was detected using affinity-purified SAD-specific polyclonal antibody and peroxidase-linked secondary antibody with the ECL kit (Amersham Life Science). Hexahistidine tagged protein was also detected using RGS•His Antibody (QIAGEN Inc.). Phosphotyrosine-containing proteins were detected by Western blotting with monoclonal antibody 4G10 (Upstate Biotechnology) with 3% BSA as the blocking agent.

Affinity purified antipeptide antibody raised against the C-terminus of SAD (see Example 8) recognized a specific ~55 kDa protein in transfected 293 cells with greater than 90% of the expressed protein being RIPA-insoluble. This molecular weight is consistent with the molecular weight predicted based on SAD's primary amino acid sequence (54,510 kD). SAD-transfected cells contain a prominent approximately 55 kDa tyrosine phosphorylated protein that is absent in vector controls. The

phosphorylated protein is most likely SAD itself because the band is clearly detected in IP-Westerns using anti-SAD crosslinked to protein A beads and 4G10 as the blotting antibody although anti-SAD only inefficiently immunoprecipitates.

Expression of Recombinant SAD in Schizosaccharomyces Pombe

S. pombe was used to express recombinant SAD as an approach to studying its function and regulation since this expression system has proven useful for studying Src family members (Superti-Furga et al., EMBO J. 12, 2625-2634; Superti-Furga et al., Nature Biotech. 14, 600-605). S. pombe strain SP200 (h-s leu1.32 ura4 ade210) was grown as described and transformations with pRSP expression plasmids were done by the lithium acetate method (Moreno et al., 1991; Superti-Furga et al., EMBO J. 12, 2625-2634). Cells were grown in the presence of 1 uM thiamine to repress expression from the nmt1 promoter or in the absence of thiamine to induce expression.

Under derepressing conditions, SAD-expressing strains show no growth defect compared to vector controls in either liquid culture or solid media. This result contrasts with the toxicity caused by expression of several other tyrosine kinases including Src and Frk. SAD protein can be detected in these strains as a weak band on Western blots using the polyclonal antibody against the C-terminus. On anti-phosphotyrosine Western blots, SAD itself is the only detectable phosphotyrosine-containing protein, however in the presence of pervanadate, cellular proteins are also phosphorylated. This observation contrasts with the results seen for Src and MKK3 which phosphorylate many yeast proteins even in the absence of phosphatase inhibitors. These findings suggest that SAD exhibits relatively limited substrate specificity and

autophosphorylates itself. These results are consistent with the transient expression experiments in 293 cells.

Example 10: Assay for SAD Kinase Activity

The example below describes an in vitro assay for SAD kinase activity. The assay is useful for the identification of modulators of SAD activity.

Materials And Methods

S. pombe expressing hexahistidine-tagged SAD were harvested by centrifugation and lysed by the glass bead method (Superti-Furga et al., EMBO J. 12, 2625-2634) in NP-40 lysis buffer (50 mM Tris-Cl pH 7.5, 150 mM NaCl, 1% NP-40, 5 mM 2-mercaptoethanol , 1 mM sodium vanadate, 1 mM phenylmethylsulfonyl fluoride, 2 µg/mL aprotinin, 1 µg/mL leupeptin, and 25 µg/mL trypsin inhibitor). Immunoprecipitations were done by mixing yeast extract (100 µg total protein in 100 µL NP-40 lysis buffer) with 0.6 µg the RGS•His Antibody (QIAGEN Inc.) and 10 µL Protein A/G agarose (Upstate Biotechnology) for 3 hrs at 4 °C. IP complexes were washed four times in 1 mL lysis buffer and once in 1 mL kinase buffer (20 mM Na-HEPES pH 7.5, 10 mM MnCl2, 2 mM 2-mercaptoethanol, and 10 µM sodium vanadate). Kinase assays were for 10 min at 30 °C in 40 ul kinase buffer containing 15 µM ATP, 0.5 uCi g-32P-ATP, and either 3 µg denatured enolase or 10 µg poly-Glu-Tyr (4:1) as the substrate. Extracts were assayed using 2-10 µg total protein per reaction and IP complexes were assayed using 5 ul Protein A/G beads per assay. Reactions were terminated by the addition of SDS sample buffer and the samples were resolved on an 10% SDS polyacrylamide gel and visualized by autoradiography.

Results

SAD was able to phosphorylate both denatured enolase and poly-Glu-Tyr in vitro. Phosphorylation of both substrates was detected in crude yeast lysates expressing SAD but not in lysates from vector control strains. In addition, anti-His IP complexes from SAD-expressing strains but not control strains phosphorylated both denatured enolase and poly-Glu-Tyr.

Example 11: Isolation Of cDNA Clones Encoding PTP05 and PTP10

The example below describes the isolation and identification of new PTP sequences from primary murine fat and rat basal forebrain and the subsequent cloning of a full-length PTP05 sequence Also described are probes useful for the detection of PTP05 and/or PTP10 in cells or tissues.

Materials and Methods:

Total RNAs were isolated using the Guanidine Salts/Phenol extraction protocol of Chomczynski and Sacchi (P. Chomczynski and N. Sacchi, Anal. Biochem. 162, 156 (1987) from ob/ob mouse fat and, separately, rat basal forebrain. These RNAs were used to generate single-stranded cDNA using the Superscript Preamplification System (GIBCO BRL, Gaithersburg, MD.; Gerard, et al, FOCUS 11:66, 1989) under conditions recommended by the manufacturer. A typical reaction used 10 μg total RNA with 1.5 μg oligo$(dT)_{12-18}$ in a reaction volume of 60 μL. The product was treated with RNaseH and diluted to 100 μL with H_20. For subsequent PCR amplification, 1-4 μL of this sscDNA was used in each reaction.

Degenerate oligonucleotides were synthesized on an Applied Biosystems 394 DNA synthesizer using established

phosphoramidite chemistry, precipitated with ethanol and used unpurified for PCR. The sequence of the degenerate oligonucleotide primers follows:

PTPDFW = 5'-GAYTTYTGGVRNATGRTNTGGGA- (sense) (SEQ ID NO: 17) and

PTPHCSA = 5'-CGGCCSAYNCCNGCNSWRCARTG -3' (antisense) (SEQ ID NO: 18).

These primers were derived from the peptide sequences DFWXMXW(E/D) (SEQ ID NO: 19) (sense strand from PTP catalytic domain) and HCXAGXG (SEQ ID NO: 20) (antisense strand from PTP catalytic domain), respectively. The standard UIPAC designations for degenerate residue designations are: N = A, C, G, or T; R = A or G; Y = C or T; V = A, C or G; W = C or T; S = C or G; M = A or C; and H = A, C or T.

PCR reactions were performed using degenerate primers applied to the single-stranded cDNA listed above. The primers were added at a final concentration of 5 μM each to a mixture containing 10 mM Tris·HCl (pH8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM each deoxynucleoside triphosphate, 0.001% gelatin, 1.5 U AmpliTaq DNA Polymerase (Perkin-Elmer/Cetus), and 1-4 μL cDNA. Following 3 min denaturation at 95 °C, the cycling conditions were 94 °C for 30 sec, 50 °C for 1 min, and 72 °C for 1 min 45 sec for 35 cycles. PCR fragments migrating between 350-400 bp were isolated from 2% agarose gels using the GeneClean Kit (Bio101), and T-A cloned into the pCRII vector (Invitrogen Corp. U.S.A.) according to the manufacturer's protocol.

Colonies were selected for mini-plasmid DNA-preparations using Qiagen columns and the plasmid DNA was sequenced using cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products were run on an ABI Prism 377 DNA Sequencer, and analyzed using the BLAST alignment algorithm (Altschul, S.F. et

al., J. Mol. Biol.215:403-10). Several copies of a clone encoding a novel PTP (R90-2-22), designated SuPTP05, was isolated from murine adipose tissue. A related clone, PTP10, was isolated from rat basal forebrain.

To obtain full-length cDNA encoding the novel phosphatase PTP05, RACE (rapid amplification of cDNA ends) was performed with sense or anti-sense oligonucleoides derived from the original PCR fragments. Marathon-Ready cDNA (Clontech, Palo Alto, CA) made from mouse testis was used in the RACE reactions with the following primers:

RACE primers:
5'-CACCGTTCGAGTATTTCAGATTGTGAAGAAGTCC-3' (6595) (SEQ ID NO:21),
5'-GGACTTCTTCACAATCTGAAATACTCGAACGGTG-3' (6596) (SEQ ID NO:22),
5'-CCGTTATGTGAGGAAGAGCCACATTACAGGACC-3' (6599) (SEQ ID NO:23),
5'-GGTCCTGTAATGTGGCTCTTCCTCACATAACGG-3' (6600) (SEQ ID NO:24),
AP-1, and AP-2 (Clontech).

RT-PCR primers for PTP05 sequeqncing:
5'-CACCGTTCGAGTATTTCAGATTGTGAAGAAGTCC-3' (6595) (SEQ ID NO:21),
5'-GGTCCTGTAATGTGGCTCTTCCTCACATAACGG-3' (6600) (SEQ ID NO:24).

Isolated cDNA fragments encoding SuPTP05 were confirmed by DNA sequening and subsequently used as probes for the screening of a murine testis cDNA library.

Two murine testis cDNA libraries (lZapII, Stratagene, La Jolla, CA and lgt10, Clontech), were screened to isolate full-length transcripts encoding PTP05. The 5' or 3'-RACE fragments were ^{32}P-labeled by random priming and used as hybridization probes at $2x10^6$ cpm/mL following standard techniques for library screening. Pre-hybridization (3 hrs) and hybridization (over-night) were conducted at 42 °C in 5X SSC, 5 X Denhart's solution, 2.5% dextran sulfate, 50 mM Na2PO4/NaHPO4 [pH 7.0], 50% formamide with 100 mg/mL denatured salmon sperm DNA.

Stringent washes were performed at 65 °C in 0.1X SSC and 0.1% SDS. Several overlapping clones were isolated and found to span the collective sequences of the PCR fragment (R90-2-22) and the RACE products. The final sequence weas verified by sequencing of both strains using a cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products were run on an ABI Prism 377 DNA Sequencer. A full-length PTP10 clone can be obtained using the same techniques.

Results:

The primary murine PTP05 transcript is 1785 nucleotides and encodes a predicted polypeptide of 426 amino acids with a predicted molecular weight of 49122 daltons (SEQ ID NO:3 and SEQ ID NO:11). The PTP05 coding sequence is flanked by a 198 nucleotide 5'-untranslated region and a 279 nucleotide 3'-untranslated region ending with a poly(A) tail. There are inframe stop codons in all three frames upstream of the primary open reading frame. The ATG beginning at nucleotide position 199 conforms to the Kozak consensus for an initiating methionine. One clone (#6.1) containes an insertion of 111 bp at nucleotide 328 resulting in an addition 37 amino acids added inframe to the coding sequence. A second clone (#10.1) has a deletion of 93 bp beginning at nucleotide 1415, resulting in a frame-shift and premature termination. Upstream of the 198bp 5'UTR, the numerous clones diverge into 2 groups, extending the 5'UTR an additional 98-153 bp. Furthermore, one clone (#15.3) lacks the polyA tail at nucleotide 1758 extends the 3' UTR by another 300 nucleotides.

The amino acid sequence shows no signal sequence or a transmembrane domain, and PTP05 is therefore predicted to be an intracellular protein. The N-terminal domain of murine PTP05 extends from amino acid 1 to 187 and is unique, *i.e.* contains no significant homology to any protein in the non-redundant

protein database. The non-redundant protein database consists of peptide sequences from GenBank Genpept, PIR, and SwissProt. There is a single protein tyrosine phosphatase catalytic domain extending from amino acids 188-420. The catalytic domain shares a relatively low level of identity at the amino acid level (40-47%) to PTPs from 5 distinct families: ZPEP (mouse) (46.7%), PTP-BAS (human) (45.6%), DEP (human) (40.5%), PTP-g (human) (40.6%), suggesting that it represents a new family of PTPs. The C-terminal tail of PTP05 extends beyond the catalytic domain from amino acids 421-426 and is not homologous to other protein tyrosine phosphatases. Motifs found in the cytoplasmic domain of other mammalian PTPs that are absent from PTP05 include: SH2, Talin/Ezrin-like, PEST, GLGF, and Retinaldehyde-binding protein domains. Owing to its divergent catalytic domain and absence of well-known non-catalytic motifs, we have designated PTP05 as a new and distinct family of protein tyrosine phosphatases.

An alternative form of murine PTP05 contains an insertion of 111-bp in the N-terminal coding region, extending the sequence by 37 aa (SEQ ID NO:4 and SEQ ID NO:12). This 1,896 bp "long" form of murine PTP05 encodes a polypeptide of 463 amino acids with a predicted molecular weight of 53716 daltons. The insertion is located at amino acid positions 44-80 and is not significantly homologous to other proteins in the non-redundant protein database.

A third form of PTP05 has a deletion of nucletotides 1415-1507 resulting in a frame shift and C-terminal truncation leading to an alternate sequence from amino acids 406-412 (SEQ ID NO:5 and SEQ ID NO:13). The 1,692 bp "C-trunc" murine PTP05 encodes a polypeptide of 412 amino acids with a predicted molecular weight of 47233 daltons.

The rat PTP10 clone shares 92% identity at the DNA level (320 nucleotides) and 85% amino acid identity at the protein

level (107 amino acids) with murine PTP05 (See Figure 1). The level of homology of the two catalytic domains suggests that PTP05 and PTP10 are distinct but related genes, and thus PTP10 is considered to be a second member of this new PTP family. Partial sequences of rat PTP10 are shown in SEQ ID NO:6 (nucleic acid) and SEQ ID NO:14 (amino acid).

Example 12: Expression of PTP05

The example below shows the evaluation of PTP05 and PTP10 expression in normal murine tissues. A similar analysis can be done in human tissues using a human PTP05 or PTP10.

Materials and Methods:

A mouse normal tissue Northern blot containing 2 μg polyA+ mRNA per lane from 8 different mouse adult tissues (lung, testis, brain, heart, liver, kidney, spleen, skeletal muscle) on a charge-modified nylon membrane was obtained from Clontech (#7762-1, Palo Alto, CA).

The membrane was hybridized with randomly primed [a^{32}P]dCTP-labeled probe synthesized from a 241 bp EcoRI fragment of R90-2-22 (see above). Hybridization was performed at 42 °C overnight in 5X SSC, 2% SDS, 10X Denhardt's solution, 50% formamide, 100 µg/mL denatured salmon sperm DNA with 1-2 x 10^6 cpm/mL of ^{32}P-labeled DNA probe. The membrane was washed at room temperature in 2X SSC/0.05% SDS for 30 min and followed by 50 °C in 0.2X SSC/0.1% SDS for 30 min, and exposed overnight on Kodak XAR-2 film.

A similar analysis was performed using the 320 bp rat PTP10 fragment as a probe of a rat normal tissue Norther blot.

RT-PCR Detection of Novel PTPs

Total RNA was isolated from fresh frozen mouse or rat (separately) tissues by centrifugation thrugh a cesium chloride cushion. Twenty µg of total RNA was reverse transcribed with random hexamers and Moloney murine leukemia virus reverse transcriptase (Super-ScriptII, GIBCO BRL, Gaithersburg, MD). PCR was then used to amplify cDNA encoding SuPTP05. RT-PCR reactions lacking only the reverse transcriptase were performed as controls. PCR products were electrophoresed on 3% agarose gels, visualized by ethidium bromide staining and photographed on a UV light box. The intensity for a 161-bp fragment specific to murine PTP05 were compared among different RNA samples. A rating of 3 represents large quantities of PTP05 transcript identified by Northern blot analysis while a rating of 0 represents little or none of the transcript was detected.

Results:

By Northern analysis, a single murine PTP05 mRNA transcript of approximately 3.4 kb was identified, and found to be exclusively expressed in the testis. The lung, brain, heart, liver, kidney, spleen, skeletal muscle samples were negative. PTP10 hybridized to a slightly smaller band and was also found only in the testis in this analysis. Northern analysis identified two rat PTP10 mRNA transcripts of approximately 3.3 kb and 1.8 kb, exclusively expressed in the testis. The rat heart, brain, spleen, lung, liver, skeletal muscle, and kidney samples were negative.

RT-PCR with gene specific primer-pairs showed that expression of the transcripts encoding PTP05 confirmed the results from Northern analysis and also detected low levels in adipose, kidney, small intestine, and cells/tissues of hematopoietic or immune origin including spleen, thymus, lymph node, bone marrow, and peripheral blood lymphocytes). RT-PCR

with rat PTP10 gene specific primers confirmed the results from the Northern analysis, detecting a strong signal only in rat testis sscDNA and not in templates corresponding to rat skeletal muscle, heart, kidney, spleen, adrenal gland, lung, liver, intestine, uterus, spinal cord, brain, cortex and ovary.

The reletively selective expression of PTP05 in cells of hematopoetic or immune origin suggests a potential involvement in immune regulation including T and B cell survival, differentiation or co-stimulation, and/or inflammatory, immunosuppressive or autoimmune disorders. Additionally, expression in adipose tissue (also the source from which PTP05 was originally isolated) suggests a possible role in metabolic disorders such as diabetes.

Example 13: Recombinant Expression Of PTP05

The following example illustrates the contruction of vectors for expression of recombinant PTP05 and the creation of recombinant cell lines expressing PTP05. Similar vectors and recombinant cell lines can be generated using PTP10 and the techniques described herein.

Contruction of Expression Vectors

Expression constructs were generated by PCR-assisted mutagenesis in which the entire coding domain of PTP05 was tagged on its carboxy-terminal end with the hemophilus influenza hemaglutinin (HA) epitope YPYDVPDYAS (SEQ ID NO:55 (Pati, supra). This construct were introduced into two mammalian expression vectors: pLXSN (Miller, A.D. & Rosman G.J., Biotechniques 7, 980-988, 1989) for the generation of virus producing lines; and pRK5 for transient expression in mammalian cells.

Dominant negative PTP05 constructs were also made in both pLXSN and pRK5 by mutation of the invariant Cys in th

conserved His-Cys-Ser-Ala-Gly motif (SEQ ID NO:56) to an Ala by PCR mutagenesis.

The entire PTP05 open reading frame excluding the initiating methionines was generated by PCR and ligated into pGEX vector for bacterial production of GST-fusion proteins for immunization of rabbits for antibody production. This vector contains the glutathione-S-transferase coding sequence followed by a polylinker for generating recombinant fusion proteins. The GST moiety comprises the N-terminal portion of the fusion protein.

Transient Expression in Mammalian Cells

The pRK5 expression plasmids (10 µg DNA/100 mm plate) containing the HA-tagged PTP05 gene can be introduced into COS and 293 cells with lipofectamine (Gibco BRL). After 72 hours, the cells were harvested in 0.5 mL solubilization buffer (20 mM HEPES pH 7.35, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1.5 mM $MgCl_2$, 1 mM EGTA, 2 mM phenylmethylsulfonyl fluoride, 1 µg/mL aprotinin). Sample aliquots were resolved by SDS polyacrylamide gel electrophoresis (PAGE) on 15% acrylamide/0.5% bis-acrylamide gels and electrophoretically transferred to nitrocellulose. Non-specific binding was blocked by preincubating blots in Blotto (phosphate buffered saline containing 5% w/v non-fat dried milk and 0.2% v/v nonidet P-40 (Sigma)), and recombinant protein was detected using a murine Mab to the HA decapeptide tag. Alternatively, recombinant protein can be detected using various PTP05-specific antisera.

Generation of Virus Producing Cell Lines

pLXSN recombinant constructs containing the PTP05 gene were transfected into an amphotropic helper cell line PA317 using $CaCl_2$ -mediated transfection. After selection on G418,

the cells were plated on normal media without G418 (500 µg/mL). Supernatants from resistant cells were used to infect the ecotropic helper cell line GP+E86, and cells again selected on G418. Resistant cells were again taken off G418, and the supernatants harvested every 8-12 hours and pooled as virus stock. Redemann et al., 1992, Mol. Cell. Biol. 12: 491-498. Viral stock titers were typically ~10^6/mL.

Stable Expression in Mammalian Cells

NIH-3T3, and BALB/3T3 cells were grown in 100 mm plates with DMEM (Gibco) containing 10% fetal calf serum (FCS). The cells were superinfected with the PTP05 retrovirus by adding approximately 3 mL viral supernatant to 15 mL culture media for approximately 24 hours. Cells expressing the retroviral constructs were then selected by growth in DMEM/10% FCS supplemented with 500 µg/mL G418.

Example 14: Generation Of Anti-PTP05 Antibodies

PTP05-specific immunoreagents were raised in rabbits against a pool of three KLH-conjugated synthetic peptides corresponding to unique sequences present in human PTP04. The peptides (see below) were conjugated at the C-terminal residue with KLH.

Peptides used for immunizing rabbits:

PTP05:

peptide 433A - MSSPRKVRGKTGRDNDEEEGNSGNLNLRN (SEQ ID NO:57)

peptide 431A - SPVLSGSSRLSKDTETSVSEKELTQLAQI (SEQ ID NO:58) and

peptide 432A - WDVSDRSLRNRWNSMDSETAGPSKTVSPV (SEQ ID NO:59).

Additional immunoreagents were generated by immunizing rabbits with a purified preparation of a GST-fusion protein containing the entire coding region of PTP05. The GST-fusion protiens were produced in DH5-alpha E. coli bacteria as described in Smith, et al Gene 67:31, 1988. Bacterial protein lysates were purified on glutathione-sepharose matrix as described in Smith, et al., supra.

Example 15: Assay for PTP05 Activity

Materials and Methods:

Recombinant wild-type and dominant negative (signaling incompetant) PTP05 (see Example 13, supra) were purified from bacteria as GST-fusion proteins. Lysates were bound to a glutathione-sepharaose matrix and washed twice with 1X HNTG, followed by one wash with a buffer containing 100 mM 2-(N-morpholino)ethansulfonic acid (MES), pH 6.8, 150 mM NaCl, and 1 mM EDTA.

The assay for phosphatase activity was essentially done as described by Pei et al.(1993) using p-nitrophenolphosphate (PNPP) as a generic PTP substrate. Briefly, after the last washing step, reactions were started by adding 50 mL Assay Buffer (100 mM MES pH 6.8, 150 mM NaCl, 10 mM DTT, 2 mM EDTA, and 50 mM PNPP) to the matrix bound proteins. Samples were incubated for 20 min. at 23 °C. The reactions were terminated by mixing 40 µL of each sample with 960 µL 1 N NaOH, and the absorbance of p-nitrophenol was determined at 450 nm. To control for the presence of PTP05 in the precipitates, the precipitates were boiled in SDS sample buffer and analyzed by SDS-PAGE. The presence of PTP05 was then detected by immunoblot analysis with anti-PTP05 antibodies.

Example 16: Isolation Of cDNA Clones Encoding ALP

The example below describes the isolation and identification of a new PTP sequence from mouse tissues and the subsequent cloning of a full-length human ALP. Also described are probes useful for the detection of ALP in cells or tissues.

Materials and Methods:

Total RNAs were isolated using a commonly known guanidine salts/phenol extraction protocol from normal mouse fat and rat pituitary. Chomczynski & Sacchi, 1987, Anal. Biochem. 162: 156. These RNA extracts were used to generate single-stranded cDNA using the Superscript Pre-amplification System (GIBCO BRL, Gaithersburg, MD.; Gerard et al., 1989, FOCUS 11: 66) under conditions recommended by the manufacturer. a typical reaction used 10 µg total RNA with 1.5 µg oligo$(dT)_{12-18}$ in a reaction volume of 60 µL. The product was treated with RNaseH and diluted to 100 µL with H_2O. For subsequent PCR amplification, 1-4 µL of this sscDNA was used in each reaction.

Degenerate oligonucleotides were synthesized on an Applied Biosystems 394 DNA synthesizer using established phosphoramidite chemistry, precipitated with ethanol and used unpurified for PCR. The sequence of the degenerate oligonucleotide primers were as follows:

PTPDFW = 5'-GAYTTYTGGVRNATGRTNTGGGA-3' (SEQ ID NO:17)
PTPHCSA = 5'-CGGCCSAYNCCNGCNSWRCARTG-3' (SEQ ID NO:18)
PTPYINA = 5'-ATCCCCGGCTCTGAYTAYATHMAYGC-3' (SEQ ID NO:60)

These primers were derived from the peptide sequences DFWXMXW(E/D) (SEQ ID NO:19) (sense strand from PTP catalytic region) and HCXAGXG (SEQ ID NO:20)(antisense strand from PTP catalytic region), and IPGSDYI(N/H)A (SEQ ID NO:61) respectively. The standard UIPAC designations for degenerate residue

designations are: N = A, C, G, or T; R = A or G; Y = C or T; V = A, C or G; W = C or T; S = C or G; M = A or C; and H = A, C or T.

PCR reactions were performed using degenerate primers applied to the single-stranded cDNA listed above. The primers were added at a final concentration of 5 μM each to a mixture containing 10 mM TrisHCl (pH8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM each deoxynucleoside triphosphate, 0.001% gelatin, 1.5 U AmpliTaq DNA Polymerase (Perkin-Elmer/Cetus), and 1-4 μL cDNA. Following 3 min denaturation at 95°C, the cycling conditions were 94 °C for 30 s, 50 °C for 1 min, and 72°C for 1 min 45 s for 35 cycles. PCR fragments migrating between 350-400 bp were isolated from 2% agarose gels using the GeneClean Kit (Bio101), and T-A cloned into the pCRII vector (Invitrogen Corp. U.S.A.) according to the manufacturer's protocol.

Colonies were selected for mini plasmid DNA-preparations using Qiagen columns and the plasmid DNA was sequenced using cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products were run on an ABI Prism 377 DNA Sequencer, and analyzed using the BLAST alignment algorithm. Altschul et al., J. Mol. Biol. 215: 403-410. A single clone encoding a novel PTP (S50-151), designated murine ALP, was isolated from murine adipose tissue using degenerate oligonucleotides PTPDFW (SEQ ID NO: 17) and PTPHCSA (SEQ ID NO:18), and a related rat ALP clone was isolated from rat pituitary using degenerate oligonucleotides PTPYINA (SEQ ID NO:60) and PTPHCSA (SEQ ID NO:18).

To isolate a full-length human ALP a human cDNA library was constructed in lambda ZapII (Stratagene, La Jolla, CA) from polyA+ RNA isolated from the human neuroblastoma cell line IMR32. The library was screened to isolate full-length

transcripts encoding ALP. The murine ALP fragment was ^{32}P-labeled by random priming and used as a hybridization probe at $2x10^6$ cpm/mL following standard techniques for library screening. Pre-hybridization (3 h) and hybridization (overnight) were conducted at 42 °C in 5X SSC, 5 X Denhart's solution, 2.5% dextran sulfate, 50 mM $Na_2PO_4/NaHPO_4$ [pH 7.0], 50% formamide with 100 mg/mL denatured salmon sperm DNA. Stringent washes were performed at 65 °C in 0.1X SSC with 0.1% SDS. Multiple clones were isolated and one 4.5 kb clone spanned the entire coding region of ALP. The final sequence was verified by sequencing of both strands using a cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products were run on an ABI Prism 377 DNA Sequencer.

Results:

The 4,456 bp human ALP nucleotide sequence encodes a polypeptide of 1,274 amino acids. The amino acid sequence shows no signal sequence or a transmembrane domain and is therefore an intracellular protein. The N-terminal end extends from amino acids 1-857 and contains several putative tyrosine phosphorylation sites and a proline-rich region (30.6% prolines) from amino acids 353-777. This proline-rich region is distantly related to plant extensin proteins (30.2% amino acid identity with Zea mays extensin-like protein GB:Z34465 using Smith-Waterman alignment) and may represent a protein interaction domain as well as the site for interaction with proteins containg SH3 motifs. The C-terminal tail of ALP extends from amino acid 1097-1274 and contains a proline/serine rich region (45.6% serines plus prolines from amino acids 1101-1214) resembling a PEST motif. This region also could serve as a target for binding proteins via their SH3 motifs.

The catalytic domain extends from amino acids 858-1096 and shares 32-37% amino acid identity to PTPs from multiple subfamilies: TC-PTP (P17706: 37.1%) PTP-BAS (D21209: 32.9%), PTPα (M34668: 34.2%), PTPβ (P23467: 34.2%), PTPσ (A49104: 33.2%), PTP1B (P20417: 34.9%), suggesting that it represents a new family of PTPs. While all other cytoplasmic PTPs have their catalytic domain at either the N- or C-terminal portion of the protein, ALP has a central catalytic domain flanked by large N- and C-terminal domains. Its catalytic domain conserves most of the invariant residues present in other PTPs, but does has several atypical amino acids. In ALP, the amino acid sequence HCSAG (SEQ ID NO:56), is changed to HCSSG (amino acid positions 1029-1033) (SEQ ID NO:75). This motif is in the catalytic site of the crystal stucture of PTP1B and PTPa, and the Ala to Ser change may effect catalyitic activity or specificty. ALP also has a change from WPD to WPE (amino acids positions 993 - 995) in its predicted surface loop of the catalytic domain. In PTP1B this Aspartate participates in a salt bridge and falls into the catalytic site on binding to a specific peptide substrate. This Asp to Glu alteration is also present in three other mammalian PTPs (PTPD1, PCP2, PTPS31).

Example 17: Expression Of ALP

The example below shows the evaluation of ALP expression in normal human tissues and in a wide variety of cancers.

Materials and Methods:

Northern blots were prepared by running 20 μg total RNA per lane isolated from 60 different tumor cell lines (HOP-92, EKVX, NCI-H23, NCI-H226, NCI-H322M, NCI-H460, NCI-H522, A549, HOP-62, OVCAR-3, OVCAR-4, OVCAR-5, OVCAR-8, IGROV1, SK-OV-3, SNB-19, SNB-75, U251, SF-268, SF-295, SF-539, CCRF-CEM, K-562,

MOLT-4, HL-60, RPMI 8226, SR, DU-145, PC-3, HT-29, HCC-2998, HCT-116, SW620, Colo 205, HTC15, KM-12, UO-31, SN12C, A498, CaKi1, RXF-393, ACHN, 786-0, TK-10, LOX IMVI, Malme-3M, SK-MEL-2, SK-MEL-5, SK-MEL-28, UACC-62, UACC-257, M14, MCF-7, MCF-7/ADR RES, Hs578T, MDA-MB-231, MDA-MB-435, MDA-N, BT-549, T47D). (obtained from Nick Scuidero, National Cancer Institute, Developmental Therapeutics Program, Rockville, MD). The total RNA samples were run on a denaturing formaldehyde 1% agarose gel and transferred onto a nitrocellulose membrane (BioRad, CA). Additional human normal tissue Northern blots containing 2 μg polyA+ mRNA per lane from 16 different human normal tissues (thymus, lung, colon, testis, brain, heart, liver, pancreas, kidney, spleen, uterus, prostate, skeletal muscle, PBLs, placenta, small intestine) on charge-modified nylon membranes (multiple tissue blots #7760-1 and #7766-1, Clontech, Palo Alto, CA) were also hybidized.

Nitrocellulose membranes for the total RNA samples were hybridized with randomly primed [gamma-^{32}P]dCTP-labeled probes synthesized from a 1 kb fragment of EcoRI-NotI of ALP. Hybridization was performed overnight at 42 °C in 4X SSPE, 2.5X Denhardt's solution, 50% formamide, 200 μg/mL denatured salmon sperm DNA, 100 μg/mL yeast tRNA (Boehringer Mannheim,IN), 0.2% SDS with 5 x 10^6 cpm/mL of [gamma-^{32}P]dCTP-labeled DNA probe on a Techne Hybridizer H-1. The blots were washed with 2X SSC, 0.1% SDS, at 65 °C for 20 min twice followed by 0.5 X SSC in 0.1% SDS at 65 °C for 20 min. The blots were exposed to a phospho-imaging screen for 24 hours and scanned on a Molecular Dynamics Phosphoimager SF.

For Clontech nylon-membrane blots, hybridization was performed at 42 °C overnight in 5X SSC, 2% SDS, 10X Denhardt's solution, 50% formamide, 100 μg/mL denatured salmon sperm DNA with 1-2 x 10^6 cpm/mL of [gamma-^{32}P]dCTP-labeled DNA probe. The

blots were washed at room temperature in 2X SSC/0.05% SDS for 30 min and followed by at 50 °C in 0.2X SSC/0.1% SDS for 30 min, and exposed for 48 hours on Kodak XAR-2 film.

For analysis of expression using reverse-transcriptase-PCR detection, total RNA was isolated from various cell lines or fresh frozen tissues by centrifugation through a cesium chloride cushion. 20 µg of total RNA was reverse transcribed with random hexamers and Moloney human leukemia virus reverse transcriptase (Super-ScriptII, GIBCO BRL, Gaithersburg, MD). PCR was then used to amplify cDNA encoding ALP. Reverse transcriptase PCR (RT-PCR) reactions lacking only the reverse transcriptase were performed as controls. PCR products were electrophoresed on 3% agarose gels, visualized by ethidium bromide staining and photographed on a UV light box.

The intensity of the fragment specific to ALP were compared among different RNA samples. A rating of 4 represents large quantities of ALP transcript while a rating of 0 represents little or none of the transcript was detected. It should be noted that detection of proteins by RT-PCR indicates a relatively higher abundance than detection by Northern blot as the RT-PCR technique utilizes total RNA whereas Northern blot analysis is performed using an enriched RNA source (mRNA).

Results:

A single ALP mRNA transcript of approximately 5.0 kb was visualized by Northern analysis. This transcript was identified in most of the normal tissue samples tested. However, the Northern analysis results shown in the Table 1 illustrate that the relative abundance of ALP mRNA is quite divergent. In normal tissues, ALP was identified in highest quantities in pancreas, followed by heart, testis, and skeletal muscle. Lower levels of the ALP transcript were identified in placenta,

thymus, lung, brain, liver, spleen, uterus, prostate and small intestine. None of the ALP transcript was detected in colon, kidney and peripheral blood leucocytes (PBLs). ALP expression was also detected in normal human adipocytes by RT-PCR methods.

In Northern blots of total RNA from human tumor cell lines, the ALP RNA transcript was most abundant in NCI-H226 (lung tumor), SK-OV-3 (ovarian tumor), and RPMI 8226 (leukemia) cell lines. The transcript was identified at lower amounts in SNB-19 (CNS tumor), SF-268 (CNS tumor), SN12C (kidney tumor), SK-MEL-2 (melanoma), UACC-62 (melanoma), and UACC-257 (melanoma) cell lines. The ALP transcript was not detected in the remaining of 44 human tumor cell lines. A summary of expression of ALP is shown in Table 1 below.

Table 1

Cell type	Origin	ALP
Thymus	Normal tissue	0.5*
Lung	Normal tissue	0.5*
Colon	Normal tissue	0*
Testis	Normal tissue	2*
Brain	Normal tissue	0.5*
Heart	Normal tissue	2*
Liver	Normal tissue	0.5*
Pancreas	Normal tissue	3*
Kidney	Normal tissue	0*
Spleen	Normal tissue	0.5*
Uterus	Normal tissue	0.5*
Prostate	Normal tissue	0.5*
Skeletal muscle	Normal tissue	2*
PBLs	Normal tissue	0*
Placenta	Normal tissue	1*
Small intestine	Normal tissue	0.5*
NCI-H226	Lung tumor	4
SK-OV-3	Ovarian tumor	3
SNB-19	CNS tumor	2
U251	CNS tumor	1
SF-268	CNS tumor	2
RPMI 8226	Leukemia	3

Cell type	Origin	ALP
HTC15	Colon tumor	1
UO-31	Colon tumor	1
SN12C	Kidney tumor	2
SK-MEL-2	Melanoma	2
SK-MEL-28	Melanoma	1
UACC-62	Melanoma	2
UACC-257	Melanoma	2
T47D	Breast tumor	1

* mRNA Northern blot.

ALP exhibits increased expression in tumor cells compared to their normal tissue counterparts. This differential expression suggests a possible disregulation or involvement of ALP in development or maintenance of the transformed phenotype.

Example 18: Recombinant Expression of ALP

The following example illustrates the contruction of vectors for expression of recombinant ALP and the creation of recombinant cell lines expressing ALP.

Contruction of Expression Vectors

Expression constructs were generated by PCR-assisted mutagenesis in which the entire coding regions of ALP was introduced into the mammalian expression vectors pcDNAIII (Invitrogen) for transient expression analysis. Additional ALP constructs were made by oligonucleotide based PCR mutagenesis to convert atypical residues in the PTP-related domain back to the amino acids more commonly present in other catalytically active PTPs. These changes include: His to Tyr at amino acid 861 (See SEQ. ID. NO.:2); Ala to Gly at amino acid 902; Phe to trp at amino acid 941; Glu to Asp at amino acid 995; and Ser to Ala at amino acid 1032. Additional constructs containing paired mutations as above were generated for amino acid positions 941/1032 and 902/1032. These constructs were ligated into the pcDNAIII mammalian expression vector behind the CMV promoter.

The entire ALP open reading frame excluding the initiating methionines was generated by PCR and ligated into pGEX vector (Pharmacia Biotech, Upsala, Sweden) for bacterial production of GST-fusion proteins for immunization of rabbits for antibody production. This vector contains the glutathione-S-transferase coding sequence followed by a polylinker for generating recombinant fusion proteins. The GST moiety comprises the N-

terminal portion of the fusion protein. The various ALP mutants were also inseted into the pGEX vecotr for production of recombinant protein reagents.

Transient Expression in Mammalian Cells

The pcDNAIII expression plasmids (10 μg DNA/100 mm plate) containing the wild-type and mutant forms of the ALP gene were introduced into 293 cells with lipofectamine (Gibco BRL). After 72 hours, the cells were harvested in 0.5 mL solubilization buffer (20 mM HEPES pH7.35, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1.5 mM $MgCl_2$, 1 mM EGTA, 2 mM phenylmethylsulfonyl fluoride, 1 &g/mL aprotinin). Sample aliquots were resolved by SDS polyacrylamide gel electrophoresis (PAGE) on 15%acrylamide/0.5% bis-acrylamide gels and electrophoretically transferred to nitrocellulose. Non-specific binding was blocked by preincubating blots in Blotto (phosphate buffered saline containing 5% w/v non-fat dried milk and 0.2% v/v nonidet P-40 (Sigma)), and recombinant protein was detected using antisera specific to the amino-terminal 352 residues (see below). Recombinant ALP protein migrated approximately 180 kDa, consistent with the predicted molecular weight of the 1274 amino acid protein.

Endogenous ALP was detected as a 200 kD protein in Westerr blots of lysates from a variety of tumor cell lines including human glioblastomas (U87MG, ATCC HTB 14; U118MG, ATCC HTB 15; U138MG, ATCC HTB 16; A172, ATCC CRL 1620; Hs683, ATCC HTB 138), rodent gliomas (C6, ATCC 107), rodent pituitary tumors (ATT20, ATCC CCL 89; GH3, ATCC CCL 82.1), human neuroblastomas (SKNMC, ATCC HTB 10; IMR 32, ATCC CCL 127), and rodent adrenal pheochromocytomas (PC12, ATCC CRL 1721). ALP protein could not be immunoprecipitated from the non-transformed cell line NIH 3T3 (ATCC CRL 1658).

It is unclear why native ALP protein appears to be larger (200 kDa) than recombinant ALP detected in transfected 293 cells (180 kDa). The difference could be the result of alternative RNA splicing, or a post-translational modification in the cell lines where it is endogenously expressed. Preliminary experiments indicate that ALP is phosphorylated on serine and threonine residues in transfected 293 cells. In addition, several tyrosine-phosphorylated proteins are associated with ALP since they are detected in Western blots using an anti-phosphotyrosine antibody following immunoprecipitation of endogenous ALP from human tumor cell lines such as IMR32 after treatments with the phosphatase inhibitor pervanadate.

Generation Of Virus Producing Cell Lines

pLXSN recombinant constructs containing the ALP gene are transfected into an amphotropic helper cell line PA317 using $CaCl_2$ mediated transfection. After selection on G418, the cells are plated on normal media without G418 (500 µg/mL). Supernatants from resistant cells are used to infect the ecotropic helper cell line GP+E86, and cells again selected on G418. Resistant cells are again taken off G418, and the supernatants harvested every 8-12 hours and pooled as virus stock. Redemann et al., 1992, Mol. Cell. Biol. 12: 491-498. Viral stock titers are typically ~10^6/mL.

Stable Expression In Mammalian Cells

NIH-3T3, BALB/3T3 or other suitable cells are grown in 100 mm plates with DMEM (Gibco) containing 10% fetal calf serum (FCS). The cells are superinfected with the ALP retrovirus by adding approximately 3 mL viral supernatant to 15 mL culture media for approximately 24 hours. Cells expressing the

retroviral constructs are then selected by growth in DMEM/10% FCS supplemented with 500 µg/mL G418.

Example 19: Generation Of Anti-Alp Antibodies

ALP-specific immunoreagents were generated by immunizing rabbits with the bacterially expressed N-terminal 352 amino acid portion of ALP expressed as a GST-fusion protein. Fusion protein was affinity purified using glutathione-sepharos colums (Pharmacia). Polyclonal anti-serum against the N-terminal portion of ALP was generated by repeatedly immunizin rabbits with the purified GST-futions protein. Affinity purified ALP antibody was obtained by binding serum IgG to ALP GST-fusion protein immobilized on glutathione-sepharose an eluting with low pH and high salt.

Example 20: Assay For ALP Activity Assay For Modulators Of Catalytic Activity

Materials And Methods:

Recombinant wild-type and mutant ALP proteins are purifie from bacteria as GST-fusion proteins. Lysates are bound to glutathione-sepharose matrix and eluted with glutathione. Th purified proteins are then washed with 2 x 1 mL HNTG, followe by one wash with 1 mL of a buffer containing 100 mM 2-(N morpholino)ethansulfonic acid (MES), pH 6.8, 150 mM NaCl, and mM EDTA. The assay for phosphatase activity is essentiall done as described by Pei et al.(1993) using p nitrophenolphosphate (PNPP) as a generic PTP substrate Briefly, after the last washing step, reactions are started b adding 50 mL Assay Buffer (100 mM MES pH 6.8, 150 mM NaCl, 1 mM DTT, 2 mM EDTA, and 50 mM p-nitrophenylphosphate) to th precipitates. Samples are incubated for 20 min. at 23 °C. Th

reactions are terminated by mixing 40 µL of each sample (without beads) with 960 µL 1 N NaOH, and the absorbance of p-nitrophenol was determined at 450 nm. To control for the presence of ALP in the precipitates, the precipitates are boiled in SDS sample buffer and analyzed by SDS-PAGE. The presence of ALP is then detected by immunoblot analysis with anti-ALP antibodies.

Example 21: A Consistent Method For Determination Of ZAP70 Kinase Activity.

The following protocol describes the reagents and procedures used to determine Zap70 protein kinase activities measuring phosphorylation of Band III-GST as readout. This assay is used in search for inhibitors of Zap70.

Materials and Reagents

1. Baculovirus (Pharmingen, CA) encoding for mutationally activated form of Zap70, in which a tyrosine residue at position 492 is replaced with a phenylalanine residue (Y492F), containing a C-terminal HA tag and a N-terminal GST tag (GST-Zap70-HA) is used. The modified protein is termed GZH (*i.e.* Y492F GST-Zap70-HA = GZH).

2. Cell lysates: SF9 cells were infected with the GZH virus at MOI of 10 for 96 hours. The cells were then washed once with PBS and lysed in lysis buffer. Insoluble material was removed by centrifugation (5 min. at 10 000 x g). Aliquots of lysates were frozen in dry ice/ethanol and stored at -80 °C until use.

3. Band III-GST: Band III-GST fusion protein (amino acid sequence: MEELQDYEDMMEEN (SEQ ID NO:62)) was expressed in XL1 Blue cells transformed with pGEX -2TK-Band III. Protein

expression was induced by addition of 0.5 mM IPTG while shaking the bacterial culture for 18 hours at 25 °C. Band III-GST by was purified by Glutathione affinity chromatography, Pharmacia, Alameda, CA

4. Biotinylated ITAM peptide 242 (ZETA-pY), Sequence: YQQGQNQLpYNELNLGRREEpYDVLDKRRGRD (SEQ ID NO:63) (Protein Chemistry Laboratory, SUGEN, INC., Redwood City, CA).

5. DMSO, Sigma, St. Louis, MO

6. 96 Well ELISA Plate: Corning 96 Well Easy Wash, Modified Flat Bottom Plate. Catalog # 25805-96.

7. NUNC 96-well V-bottom polypropylene plates for dilution of compounds. Applied Scientific Catalog No. AS-72092

8. Streptavidin: Sigma S-8276

9. Purified Rabbit anti-GST antiserum. AMRAD catalog # 9001605

10. Goat anti-Rabbit-IgG-HRP. Amersham Catalog No. V010301

Buffer solutions:

Lysis buffer:	Kinase buffer:
10 mM Tris, pH 7.5	10 mM $MgCl_2$
150 mM NaCl	10 mM $MnCl_2$
1% NP40	10 mM DTT
1 mM PMSF	20 mM HEPES/Cl, pH 7.5
0.4 mM Na_3VO_4	20 mM β-glycerophosphate
2 mg/ml Leupeptin	100 mM Na_3VO_4
2 mg/ml Aprotinin	

Blocking buffer:	Wash buffer (TBST):
10 mM Tris, pH 7.5	50 mM Tris, pH 7.5
100 mM NaCl	150 mM NaCl
0.1% Tween 20	0.1% Tween 20

1% BSA

Procedure:

Preparation of Streptavidin Coated ELISA Plates:

Prepare borate buffer by titrating 0.1 M boric acid with 0.1 M sodium borate to pH 8.7. Add sodium azide to a final concentration of 0.05% and store at 4 °C. Prepare 1 mg/ml Streptavidin in borate buffer and store at 100 µL aliquots at -80 °C. Coat 0.1 µg/well Streptavidin in 100 µL of borate buffer at room temperature for 18 hours. Wash wells with 200 µL cold TBST twice. Invert the plate and blot the plate dry, cover with parafilm, and store at 4 °C for no more than one week. For longer storage, plates should be stored at -80 °C.

Preparation of phosphotyrosine antibody-coated ELISA plates:

Coat 1 µg/well 4G10 (Upstate Biotechnology, NY) in 100 µL of PBS overnight at 4 °C and block with 200 µL of blocking buffer for at least hour.

Kinase Assay Procedure

Biotinated peptide 242 was bound to the ELISA Plate by incubating 1 µg/well in 100 µL PBS overnight at 4 °C with streptavidin coated ELISA Plate (see above). The wells were blocked with 200 µL blocking buffer for 30 minutes at room temperature, after which the blocking buffer was removed by aspiration. Insect cell lysate containing the Zap70 fusion protein (GZH) was added (30 µg/well, volume adjusted to 100 µL/well with lysis buffer) and left to incubate at 4 °C for 2 hours. The lysate was removed by aspiration and the wells washed with TBST. Substrate and test compound (if any) were

added and allowed to stand for 15 minutes (GST-Band III, 5 µg/well in 90 µL final volume). The kinase reaction was started by the addition of 10 µL of 0.1 mM ATP per well for a final concentration of 10 µM. The 96 well plate was left for 30 minutes at room temperature (shaking) after which 90 µL of the reaction liquid was transferred to wells in a 96 well plate previously coated with an anti-phosphotyrosine antibody (UB40, Upstate Biotechnology, NY). This plate was allowed to stand for 30 minutes at room temperature, after which the liquid was removed and the wells washed with TBST. Rabbit anti-GST antibody was added (0.1 µg/well in 100 µL blocking buffer) and incubated for 30 minutes at room temperature. The liquid was again removed and the wells washed with TBST. Goat anti-Rabbit-IgG-HRP was added at 1:40,000 dilution in 100 µL of blocking buffer for 30 minutes at room temperature, after which it was removed and the wells washed with TBST and developed with ABTS. The plate is then read in an ELISA plate reader at 410 nm. If the protein being tested is a captured protein, the reading from the ELISA plate reader can be related to the modulating activity of the test compound when it is compared with the activity of a control protein.

Example 22: Isolation And Characterization Of ALK-7

In order to isolate ALK-7, we designed degenerate oligonucleotides encoding amino acid motifs within kinase subdomains II and VI common to all known mammalian STK receptors. (Hanks and Hunter, FASEB J. 9:576-595, 1995) Subdomain II is at the N-terminus of the kinase domain and contains the invariant lysine residue that is essential for enzyme activity and is involved in ATP binding by interacting with the a- and b-phosphates of all kinases whose structure has been elucidated. Subdomain VI

is referred to as the catalytic loop and contains the consensus motif HRDLKXXN (SEQ ID NO:64). The Asp residue is involved in accepting the proton from the hydroxyl group during the phosphotransfer process key to all protein kinases. Based on comparison of all STK receptors, we designed degenerate oligonucleotide primers to these subdomains that would recognize both type I and type II STK receptors.

When this PCR strategy was applied to a human neuroblastoma cell line (SY5Y) sscDNA as a template, multiple copies of a novel DNA fragment (ALK-7) were isolated that exhibited significant homology to other STK receptors. The novel sequence was most similar to ALK-4 (Franzen, *et al.*, Cell 75(4):681, 1993) and ALK-5 (ten Dijke, *et al.*, Oncogene 8(10):2879, 1993) and was referred to as ALK-7.

Materials And Methods

Total RNAs were isolated using the Guanidine Salts/Phenol extraction protocol of Chomczynski and Sacchi (P. Chomczynski and N. Sacchi, Anal. Biochem. 162, 156 (1987) from normal human tissues, from regional sections of human brain, from cultured human tumor cell lines, and from primary neonatal rat sympathetic, motor, and sensory neuronal cells, as well as mesothalamic dopaminergic neurons.

These RNAs were used as templates to generate single-stranded cDNAs using the Superscript Preamplification System for First Strand Synthesis kit purchased from GibcoBRL (Life Technologies, U.S.A.; Gerard, G.F. *et al.* (1989), FOCUS 11, 66) under conditions recommended by manufacturer. A typical reaction used 10 µg total RNA or 2 µg poly(A)$^+$ RNA with 1.5 µg oligo(dT)$_{12-18}$ in a reaction volume of 60 µL. The product was treated with RNaseH and diluted to 100 µL with H_2O. For

subsequent PCR amplification, 1-4 μL of these sscDNAs were used in each reaction.

Oligonucleotides were synthesized on an Applied Biosystems 394 DNA synthesizer using established phosphoramidite chemistry and were used unpurified after precipitation with ethanol. The degenerate oligonucleotide primers are:

STK1 = 5'-GARRARGT6GC6GT6AARRT6TT-3' (SEQ ID NO:65)(sense)

STK3- = 5'-TTRATRTC6CKRTG6GM6AT6GM6GGYTT-3' (SEQ ID NO:66) (antisense).

These primers were derived from the peptide sequences **E(K/E)VAVK(V/I)F** (SEQ ID NO:67) (sense strand from kinase subdomain II) and **KP(A/S)I(A/S)HRDIK** (SEQ ID NO:68)(antisense strand from kinase subdomain VI), respectively. Degenerate nucleotide residue designations are: N = A, C, G, or T; R = A or G; Y = C or T; M = A or C; K = G or T; and 6 = Inosine. Using ALK1 as a template, these primers produce a product of 321 bp.

A PCR reaction was performed using primers STK1 and STK3- applied to the single-stranded sources listed above. The primers were added at a final concentration of 5 μM each to a mixture containing 10 mM Tris HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 uM each deoxynucleoside triphosphate, 0.001% gelatin, and 1.5 U AmpliTaq DNA Polymerase (Perkin-Elmer/Cetus), and 1-4 ul cDNA. Following 3 min denaturation at 95°C, the cycling conditions were 94 °C for 30 s, 37 °C for 1 min, a 2 min ramp to 72 °C, and 72 °C for 1 min for the first 3 cycles, followed by 94 °C for 30 s, 50 °C for 1 min, and °C for 1 min 45 s for 35 cycles. PCR fragments migrating at ~320 bp were isolated from 2% agarose gels using GeneClean (Bio101), and T-A cloned into the pCRII vector (Invitrogen Corp. U.S.A.) according to the manufacturer's protocol.

Colonies were selected from mini plasmid DNA-preparations using Qiagen columns and the plasmid DNAs were sequenced using cycle sequencing dye-terminator kit with AmpliTaq DNA Polymerase, FS (ABI, Foster City, CA). Sequencing reaction products were run on an ABI Prism 377 DNA Sequencer, and analyzed using the BLAST alignment algorithm (Altschul, S.F. *et al.*, J. Mol. Biol. 215:403-10). A novel clone (STKR6.22) was isolated by PCR with primers STK1 and STK3- on single-stranded cDNA from human SY5Y cells as a template. This clone was subsequently designated as a fragment of human ALK-7.

A lambda gt11 (Clontech, Palo Alto, CA) cDNA library was constructed using mRNA from a pool of nine whole human pituitary glands. Phage were screened on nitrocellulose filters with the random primed ^{32}P-labeled insert from STKR6.22 encoding human ALK-7 at $2x10^6$ cpm/mL in hybridization buffer containing 6xSSC, 1x Denhardt's reagent, 0.1% SDS, with 0.1 mg/mL denatured, fragmented salmon sperm DNA. After overnight hybridization at 65 °C, filters were washed in 0.1xSSC, 0.1% SDS at 65 °C. Full length cDNA clones were sequenced on both strands using manual sequencing with T7 polymerase and oligonucleotide primers (Tabor and Richardson, 1987, Proc. Natl. Acad. Sci., U.S.A. 84:4767-71).

Results

Two overlapping cDNA clones (P6 and P7), spanning 1794 nucleotides were isolated from a human pituitary library. This sequence contains an ATG at position 156 that conforms to the Kozak consensus for translational initiation and is followed by a 1,482 nucleotide open reading frame with the capacity to encode a polypeptide of 493 amino acids. There are no other initiation codons 5' to the ATG located at position 156. The coding region for human ALK-7 is flanked by 5' and 3' untranslated regions of 155 and 157, respectively. There is no

polyadenylated region although the 3' end of the sequence shown in SEQ ID NO:8 is noticeably AT-rich, a feature characteristic of sequences from 3'-untranslated regions. An additional cDNA clone (P4) extended an additional 1 kb 3' of this sequence.

DNA sequence determination was performed with dideoxy terminators using Sequenase 2.0. A primer walking strategy on both strands was used to confirm the complete nucleotide sequence. Oligonucleotide primers were made with an ABI 348 DNA synthesizer.

A Smith-Waterman search with the human ALK-7 gene sequence of the public nonredundant nucleic acid and EST databases revealed no identical matching sequences confirming that this is a novel human gene. The closest match to the human ALK-7 sequence (85% nucleic acid identity) is a recent entry (GenBank ACC:U69702) which appears to be the rat orthologue of human ALK-7.

The 493 amino acid human ALK-7 sequence contains two hydrophobic regions from 1-25 and 114-138. (See SEQ ID NO:16) The first hydrophobic region meets the criteria of a signal peptide domain, with a discriminant score of 5.76 using the method of McGeoch (D. J. McGeoch, Virus Research, 3, 271, 1985), and with a weight matrix score of +6.75 (threshold = 3.5) using the von Heijne algorithm (G. von Heijne, Nucl. Acids Res., 14, 4683, 1986). The second hydrophobic region generates a likelihood score of -9.34, using the ALOM method of Klein *et al.* (P. Klein, M. Kanehisa, and C. DeLisi, Diochim. Biophys. Acta, 815, 468, 1985) to predict transmembrane domains. This algorithm predicts a maximal range of the transmembrane domain to be from aa 108-138.

Based on this analysis, ALK-7 is predicted to be a type Ia integral membrane protein with a molecular weight of 52.35 kD after cleavage of the N-terminal signal peptide.

Example 23: Expression Of ALK-7

Using both Northern blots and PCR analysis with the novel fragment originally cloned from SY5Y cells as described above as a probe, we screened RNAs using from a large number of tumor cell lines and multiple human tissues, demonstrating an apparent selectivity in expression of ALK-7 in neuronal cells from the pituitary and substantiate nigra.

Materials And Methods

Northern Blot Analysis

Northern blots were obtained from Clontech (Palo Alto, CA) containing 2 μg polyA+ RNA from 16 different adult human tissues (spleen, thymus, prostate, testis, ovary, small intestine, colonic mucosa, heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, and peripheral blood leukocytes), and four different human fetal tissues (brain, lung, liver, and kidney), on a charge-modified nylon membrane. Additional Northern blots were prepared by running 20 μg total RNA on formaldehyde 1.2% agarose gel and transferring to nylon membranes.

Filters were hybridized with random prime [^{32}P]dCTP-labeled probes synthesized from the 320 bp insert from human ALK-7 clone STKR6.22. Hybridization was performed at 60 °C overnight in 6XSSC, 0.1% SDS, 1X Denhardt's solution, 100 mg/mL denatured herring sperm DNA with 1-2 x 10^6 cpm/mL of ^{32}P-labeled DNA probes. The filters were washed in 0.1XSSC/0.1% SDS, 65 °C, and exposed overnight on Kodak XAR-2 film.

Semi-Quantitative RT-PCR Detection

The expression pattern of ALK-7 was also investigated using a PCR technique, RNA was isolated from a variety of human cell lines, fresh frozen tissues, and primary tumors as

detailed above. Single stranded cDNA was synthesized from 10 ug of each RNA as described above using the Superscript Preamplification System (GibcoBRL) These single strand templates were then used in a 35 cycle PCR reaction with two human ALK-7-specific oligonucleotides:

ALK-7a: 5'-AACTTTGGCTGGTATCTGAATATC-3' (SEQ ID NO:69), and

ALK-7b: 5'-CCTTGTGTACCAACAATCTCCATA-3' (SEQ ID NO:70).

Reaction products were electrophoresed on 2% agarose gels, stained with ethidium bromide and photographed on a UV light box. The relative intensity of the -150-bp ALK-7-specific bands were estimated for each sample. A similar pair of oligonucleotides was designed for detection of rat ALK-7:

4076: 5'-CTCCAGAGATGAGAGATCTTGG-3' (SEQ ID NO:71), and

4077: 5'-TTCCAGCCACGGTCACTATGTT-3') (SEQ ID NO:72), encompassing a -210 bp region of the rat gene.

Results

ALK-7 mRNA transcript was not detectable by Northern analysis from multiple human tissue sources, suggesting its expression is highly restricted. Using a more sensitive PCR-based detection, ALK-7 was found to be expressed in human substantia nigra, anterior pituitary, and Calu-6 lung carcinoma cell line (see below). Weak expression was found in several other locations including whole brain, cerebellum, and prostate. Multiple other normal human tissues and tumor cell lines showed no detectable ALK-7 expression.

HUMAN ALK-7 RNA EXPRESSION ANALYSIS

Medium (++)
Substantia Nigra
Anterior Pituitary
Calu-6 (Lung Ca)

Weak (+)

Brain
Posterior Pituitary
Cerebellum
Ovary
Prostate
Fetal Intestine
Duodenum
T48 (colon Ca)

Negative
IMR-32 (neuroblastoma)
SY5Y (neuroblastoma)
SK-N-SH (neuroblastoma)
SWI763 (astrocytoma)
SW1388 (astrocytoma)
U-138 (glioblastoma)
U87MG (glioblastoma)
Menirigiomas (1° tumors)
SKOV-3 (ovarian Ca)
ASPC (pancreas Ca)
CAPAN-1(pancreas Ca)
HS766T (pancreas Ca)
PANC (pancreas Ca)
HOS (osteoSarcoma)
KHOS (osteoSarcoma)
HTB227 (breast Ca)
HTB131 (breast Ca)
LS123 (colon Ca)
LS147T (colon Ca)
SkCO4 (colon Ca)
SW11E (colon Ca)
HTC15 (colon Ca)
SW403 (colon Ca)
HT29 (colon Ca)
SW627 (colon Ca)
SW948 (colon Ca)
HUVEC (h. endothelial)
Fibroblasts (Primary)
Pancreas
Testis
Thymus
Liver
Heart
Placenta
Lung
Skel. Muscle
Kidney
Spleen
Ovary
Colon
Leukocytes

In situ EXPESSION PROFILE of RAT ALK-7

The neuronal expression of ALK-7 was assessed by in situ analysis in sagittal and coronal sections from neonatal and

adult rat brains using a fragment of the extracellular domain of rat ALK-7 as a probe. This region was selected because its dissimilarity with the related ALK-4 and ALK-5. Other groups have performed in situs with the catalytic domain of rat ALK-7 demonstrating specific expression in neuronal tissues (cerebellum, hippocampus, and brainstem nuclei), kidney, testis, lung, dorsolateral and anterior prostate, and adipose tissue. However, the probe used in these studies contained an ALK-7 catalytic domain which may cross-react with the related ALK-4 and ALK-5 (77% nucleotide sequence identity with stretches of 27/29 and 25/26 bp identity to rat ALK-7) and thereby broaden the expression profile. Using a more selective ALK-7 probe our analysis revealed the more restricted expression. In sagital sections, a moderate strength granular band was visible in the CA2 and CA3 regions of the hippocampus, dentate dyrus, olfactory tubercle, dorsal outer layer of the cortex, and in a band crossing the frontal cortex area 2 from the exterior to the corpus callosum. A moderate signal was detected in the caudate putamen and thalamic nuclei. In addition, signals of moderate strength were detected in the region of the magnocellular nucleus of the lateral hypothalamus and the medial tuberal nucleus. A similar signal was observed in the region of the cuneiform nucleus on the anterior border of the cerebellum. The cerebellum was devoid of hybridizing ALK-7.

Coronal sections support the finding of expression in the CA2, CA3 region of the hippocampus, dentate gyrus, caudate putamen, and in the region underlying the exterior of the cortex. In addition, a signal of moderate strength was detected in the dorsomedial part of the ventromedial hypothalamic nucleus. A dispersed nuclei signal of lesser strength was detected in the area of the amygdalopiriform transition.

Example 24: ALK-7-Specific Antibodies

ALK-7-specific immunoreagents were raised in rabbits against KLH-conjugated synthetic peptide YRKKKRPNVEEPL (SEQ ID NO:76) from the juxtamembrane portion of the cytoplasmic domain of ALK-7. This region is unique to ALK-7 compared to other type I STK receptors, thereby allowing for the generation of ALK-7 specific antisera. The N-terminal extracellular domain of ALK-7 expressed as a GST-fusion was also used as an immunogen to raise polyclonal antibodies in rabbits and to generate monoclonal antibodies in mice using the techniques described above. These antibodies were used to localize expression of the endogenous and recombinant protein as describe below.

Example 25: Recombinant Alk-7 Expression

The following example describes the construction of vectors for transient and stable expression in mammalian cells. Expression constructs were generated to make wild type ALK-7 as well as a signaling incompetent ALK-7 (ALK-7DN) and a constitutively activated ALK-7 (ALK-7TA).

Materials and Methods

Construction of Vectors

Expression constructs were generated by PCR-assisted mutagenesis in which the entire coding domain of ALK-7 was tagged at its carboxy-terminal ends with the hemophilus influenza hemaglutinin (HA) epitope YPYDVPDYAS (SEQ ID NO:77) (Pati, Gene 114:285, 1992). This constructs were introduced into two mammalian expression vectors: pAdRSVOES-, a modified adenovirus vector for the generation of virus producing recombinant protein, and pRK5 for transient expression analysis.

Recombinant adenoviruses were generated by in vivo ligation as follows.

The transfer vector used Contains the following DNA sequences in order: The left terminal region of adenovirus type 5 encoding the packaging sequences (adenovirus type 5 nucleotides 1-454); the Rous Sarcoma Virus long terminal repeat promoter and the SV40 polyA region, isolated as an expression cassette from the plasmid pREP (Invitrogen Corporation); nucleotides 3320-5790 of the type 5 adenoviral genome; and the ori and beta-lactamase genes derived from the *E. coli* plasmid pBluescript. Two additional forms of the plasmid were generated. The first, pAdRSVlacZ, was prepared by the insertion of a double stranded synthetic oligonucleotide into the BamHI restriction site between the RSV promotor and the SV40 polyA sequence with the following nucleotide sequence (upper strand shown): 5' CTTCGAAAGCTTGAAATCGGTACCATCGATTCTAGAGTTAACTTCGAA. (SEQ ID NO: 73) The *E. coli* lacZ gene was excised from the expression plasmid pCMVb (Clontech, Inc.) with the enzyme Not I and inserted into the Not I site between the promoter and the polyA sequence. This generated a plasmid that expressed the lacZ gene, and had two BstBI restriction sites between the lacZ gene and the polyA region. The second plasmid (pAdRSVOES-) was generated by inserting a double stranded synthetic oligonucleotide into the same region as above. Its nucleotide sequence was the following: 5' CTCTAGAACGCGTTAAGGCGCGCCAATATCGATGAATTCTTCGAAGC. (SEQ ID NO:74 This plasmid allowed the introduction of exogenous cDNAs into the plasmid for expression purposes.

The viral DNA used for generation of recombinant viruses was derived from a virus (AdlacZBstBI) in which the left end of the adenovirus genome has been replaced by the homologous region of pAdRSVlacZ. To achieve this, DNA ,vas isolated from

the Ad5 dl327 strain of adenovirus (Jones and Shenk, Cell, 1978) (deleted in the E3 region), cleaved with ClaI enzyme, and cotransfected into the HEK2934 cell line via calcium phosphate coprecipitation with the pAdRSVlacZ plasmid. Recombinant adenovirus plaques resulting from this transfection were screened for the ability to express the lacZ gene by histochemical staining with X-Gal. The resulting recombinant adenovirus, AdlacZBstBI, provided the backbone for additional adenovirus constructs, allowing a screen for recombinant plaques based on the presence or absence of lacZ activity in that further recombination would replace the lacZ gene with the cotransfected cDNA. To achieve this, the transfer vector construct is linearized by digestion with BstBI, and cotransfected with AdlacZBstBI DNA which has also been cleaved with BstBI. Typically, 5 mg of transfer vector plasmid DNA are corecipitated with 2 mg of viral DNA for the transfection; *in vivo* ligation of viral DNA and linearized transfer vector produces a novel recombinant virus directing expression of the new transgene.

A signaling incompetent ALK-7 construct was also made in both vectors pAdRSVOES- and pRK5 by insertion of an HA-tag at aa 230 in the ALK-7 coding region just after catalytic domain II. Truncation of other Type I STKRs in an analogous location has functioned in a dominant negative manner. This construct was called ALK-7DN. A constitutively active form of ALK-7 was generated by a Thr to Asp mutation at amino acid 194 just upstream of the catalytic domain I GXGXXG motif. In other Type I STKRs, this residue undergoes ligand-dependent trans-phosphorylation by the associated Type II STKR, resulting receptor activation and initiation of a signaling cascade. A similar mutation in other Type I STKR's results in a ligand-independent, constitutively activated receptor. This construct was called ALK-7TD.

Generation Of Recombinant ALK-7 - Adenovirus

Early passage HEK293 cells (Graham, *et al.*, J. Gen. Virol. 36:59, 1977) were maintained in Dulbecco's modified Eagles medium + 10% calf serum. HEK293 monolayers were transfected with the ALK-7-encoding transfer vectors and cultured from five to seven days to allow plaques to appear. The monolayers were then stained with 25 mg/mL 5-bromo-4-chloro73-indolyl-b-D-galactopyranoside for several hours to identify non-recombinant (blue-stained) plaques. Putative recombinant plaques were screened for expression of the transgene by infection of HEK293 cultures followed by immunohistochemistry with the monoclonal antibody recognizing the HA epitope. Viruses which were positive for transgene protein expression were picked and subjected to several rounds of claque purification prior to amplification and purification on cesium chloride gradients. Banded viruses were diluted five-fold with dilution buffer (Curiel *et al.*, Proc. Natl. Acad. Sci., USA 88:8850-8854, 1991) and stored at -80 °C. Approximate titers of the virus preparations were determined immunohistochemically on HEK293 cultures. The following viruses were generated: AdRSVALK-7-HA; AdRSVALK-7-DN; and AdRSVALK-7-TD.

Transient Expression

The pRK5 expression plasmids (10 µg DNA/100 mm plate) containing the KA-tagged ALK-7, the ALK-7DN, and ALK-7TD constructs were introduced into COS and 293 cells with lipofectamine (Gibco BRL). After 72 hours, the cells were harvested in 0.5 ml solubilization buffer (20 mM HEPES pH 7.35 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1.5 mM $MgCl_2$, 1 mM EGTA, 2 mM phenylmethylsulfonyl fluoride, 1 µg/mL aprotinin)

Sample aliquots were resolved by SDS polyacrylamide gel electrophoresis (PAGE) on 15% acrylamide/0.5% bis-acrylamide gels and electroplicretically transferred to nitrocellulose. Non-specific binding was blocked by preincubating blots in Blotto (phosphate buffered saline containing 5% w/v non-fat dried milk and 0.2% v/v nonidet P-40 (Sigma)), and recombinant protein was detected using a murine Mab to the HA decapeptide tag. Alternatively, recombinant protein can be detected using various ALK-7-specific antisera.

Expression In Neuronal Cells

The recombinant ALK-7 protein described above were expressed in PC12 cells and primary rat neuronal cultures by adenovirus mediated infection. These cells will allow further investigation into ALK-7 function. Recombinant protein expression was confirmed by immunostaining with an anti-HA antibody.

PC12 cultures (Greene, *et al.*, Methods Enzymol. 147:207, 1987) were maintained in RPMI medium containing 10% horse serum and 5% fetal calf serum. Four differentiation experiments the medium was changed to RPMI containing 1X N2 supplement and 0.1% BSA, and the cells were grown on a collagen I substrate. For PC12 cell survival, the cells were grown in RPMI containing 0.1% BSA. All cultures also contained 1X penicillin/streptomycin. For adenoviral infections, PC12 cells were incubated overnight with recombinant viruses at a multiplicity of infection (MOI) between 1 and 10. The cells were then washed and replated either into differentiation or survival conditions for two days. Nerve Growth Factor (50 ng/mL) served as a positive control. For differentiation, the cultures were fixed with 2% paraformaldehyde and the percentage of cells bearing processes longer than 1 cell diameter was determined. For survival, the cultures were incubated with 0.05% MTT for

1.5 hours to stain living cells, and the relative number of cells surviving in each condition was determined.

Sympathetic and sensory neurons were isolated as described (Hawrot and Patterson, Methods Enzymol. 53:574, 1979; Fields *et al.*, Cell 14:43, 1978) and cultured in a defined medium (Hawrot and Patterson, *supra*). Sympathetic neurons were isolated from superior cervical ganglia dissected from E20 - E21 rat fetuses, while dorsal root ganglion sensory neurons were obtained from E16 - E18 rats. The ganglia were treated with 0.25% trypsin for 10 minutes, washed, and triturated to obtain a single cell suspension. Sensory neurons were preplated for 1 hour on tissue culture plastic to deplete adherent cells. Dopaminergic neurons were isolated as described (Shimoda, *et al.*, Brain Research 586:319-331, 1992) and cultured in Neurobasal medium, supplemented with B27 supplements (Life Technologies). Neurons were infected with adenoviruses for two hours on collagen I-coated tissue culture plastic (supplemented with NGF for sensory and sympathetic neurons), and the cells were then washed and allowed to recover for two to four additional hours (with NGF if appropriate). After the recovery period, the cells were washed extensively to remove the growth factor, and plated onto polylysine-laminin coated chamber slides. The addition of NGF at 50 ng/mL served as a positive control for survival of sensory and sympathetic neurons. After an additional two days to three days, the sensory and sympathetic cultures were stained with calcein AM (1 mg/mL) for 45 minutes, mounted and examined by immunofluorescence. Generally, five disperse fields representing 7% of the well were photographed and the number of surviving neurons quantitated. To determine dopaminergic neuron survival, the cultures were fixed and the number of tyrosine hydroxylase positive neurons was determined.

Results

Recombinant ALK-7 protein expressed in COS cells migrated with apparent Mr of 52kD-63kD, consistent with its predicted molecular weight of 54kD based on its primary amino acid sequence and the presence of multiple glycosylation sites. The ALK-7TD constitutive active form produced proteins indistinguishable from the wild type construct on SDS-PAGE. The ALK-7DN construct expressed proteins of Mr 23.5 kd, 28 kD and 32 kD consistent with the presence of varying amounts of glycosylation on this truncated receptor. This analysis confirms the recombinant protein can be stably produced in mammalian cells.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The molecular complexes and the methods, procedures, treatments, molecules, specific compounds described herein are presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the

terms "comprising", "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. For example, if X is described as selected from the group consisting of bromine, chlorine, and iodine, claims for X being bromine and claims for X being bromine and chlorine are fully described.

In view of the degeneracy of the genetic code, other combinations of nucleic acids also encode the claimed peptides and proteins of the invention. For example, all four nucleic acid sequences GCT, GCC, GCA, and GCG encode the amino acide alanine. Therefore, if for an amino acid there exists an average of three codons, a polypeptide of 100 amino acids in length will, on average, be encoded by 3^{100}, or 5×10^{47}, nucleic acid sequences. It is understood by those skilled in the art that, with, Thus, a nucleic acid sequence can be modified to form a second nucleic acid sequence, encoding the same

polypeptide as endoded by the first second nucleic acid sequences, using routine procedures and without undue experimentation. Thus, all possible nucleic acids that encode the claimed peptides and proteins are also fully described herein, as if all were written out in full taking into account the codon usage, especially that preferred in humans.

Furthermore, changes in the amino acid sequences of polypeptides, or in the corresponding nucleic acid sequence encoding such polypeptide, may be designed or selected to take place in an area of the sequence where the significant activity of the polypeptide remains unchanged. For example, an amino acid change may take place within a β-turn, away from the active site of the polypeptide. Also changes such as deletions (*e.g.* removal of a segment of the polypeptide, or in the corresponding nucleic acid sequence encoding such polypeptide, which does not affect the active site) and additions (*e.g.* addition of more peptides to the polypeptide sequence without affecting the function of the active site, such as the formation of GST-fusion proteins, or additions in the corresponding nucleic acid sequence encoding such polypeptide without affecting the function of the active site) are also within the scope of the present invention. Such changes to the polypeptides can be performed by those with ordinary skill in the art using routine procedures and without undue experimentation. Thus, all possible nucleic and/or amino acid sequences that can readily be determined not to affect a significant activity of the peptide or protein of the invention are also fully described herein.

Other embodiments are within the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: SUGEN, INC.
351 Galveston Drive
Redwood City, CA 94063
U.S.A.

(ii) TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF TYROSINE PHOSPHATASE-RELATED DISORDERS AND RELATED METHODS

(iii) NUMBER OF SEQUENCES: 76

(iv) CORRESPONDENCE ADDRESS:

(A)	ADDRESSEE:	Lyon & Lyon
(B)	STREET:	633 West Fifth Street Suite 4700
(C)	CITY:	Los Angeles
(D)	STATE:	California
(E)	COUNTRY:	U.S.A.
(F)	ZIP:	90071-2066

(v) COMPUTER READABLE FORM:

(A)	MEDIUM TYPE:	3.5" Diskette, 1.44 Mb storage
(B)	COMPUTER:	IBM Compatible
(C)	OPERATING SYSTEM:	IBM P.C. DOS 5.0
(D)	SOFTWARE:	FastSEQ for Windows 2.0

(vi) CURRENT APPLICATION DATA:

(A)	APPLICATION NUMBER:	To be assigned
(B)	FILING DATE:	Herewith
(C)	CLASSIFICATION:	

(vii) PRIOR APPLICATION DATA:

(A)	APPLICATION NUMBER:	US 60/044,428
(B)	FILING DATE:	April 28, 1997
(A)	APPLICATION NUMBER:	US 60/047,222
(B)	FILING DATE:	May 20, 1997
(A)	APPLICATION NUMBER:	US 60/049,477
(B)	FILING DATE:	June 12, 1997
(A)	APPLICATION NUMBER:	US 60/049,756
(B)	FILING DATE:	June 12, 1997
(A)	APPLICATION NUMBER:	US 60/049,914
(B)	FILING DATE:	June 18, 1997
(A)	APPLICATION NUMBER:	US 60/063,595
(B)	FILING DATE:	October 23, 1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Warburg, Richard J.
(B) REGISTRATION NUMBER: 32,327
(C) REFERENCE/DOCKET NUMBER: 233/032-PCT

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (213) 489-1600
(B) TELEFAX: (213) 955-0440
(C) TELEX: 67-3510

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3580 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCCGGGTGCC CTCCCTCAAC CTACTTATAG ACTATTTTTC TTGCTCTGCA GCATGGACCA 60
AAGAGAAATT CTGCAGAAGT TCCTGGATGA GGCCCAAAGC AAGAAAATTA CTAAAGAGGA 120
GTTTGCCAAT GAATTTCTGA AGCTGAAAAG GCAATCTACC AAGTACAAGG CAGACAAAAC 180
CTATCCTACA ACTGTGGCTG AGAAGCCCAA GAATATCAAG AAAAACAGAT ATAAGGATAT 240
TTTGCCCTAT GATTATAGCC GGGTAGAACT ATCCCTGATA ACCTCTGATG AGGATTCCAG 300
CTACATCAAT GCCAACTTCA TTAAGGGAGT TTATGGACCC AAGGCTTATA TTGCCACCCA 360
GGGTCCTTTA TCTACAACCC TCCTGGACTT CTGGAGGATG ATTTGGGAAT ATAGTGTCCT 420
TATCATTGTT ATGGCATGCA TGGAGTATGA AATGGGAAAG AAAAAGTGTG AGCGCTACTG 480
GGCTGAGCCA GGAGAGATGC AGCTGGAATT TGGCCCTTTC TCTGTATCCT GTGAAGCTGA 540
AAAAAGGAAA TCTGATTATA TAATCAGGAC TCTAAAAGTT AAGTTCAATA GTGAAACTCG 600
AACTATCTAC CAGTTTCATT ACAAGAATTG GCCAGACCAT GATGTACCTT CATCTATAGA 660
CCCTATTCTT GAGCTCATCT GGGATGTACG TTGTTACCAA GAGGATGACA GTGTTCCCAT 720
ATGCATTCAC TGCAGTGCTG GCTGTGGAAG GACTGGTGTT ATTTGTGCTA TTGATTATAC 780
ATGGATGTTG CTAAAAGATG GGATAATTCC TGAGAACTTC AGTGTTTTCA GTTTGATCCG 840
GGAAATGCGG ACACAGAGGC CTTCATTAGT TCAAACGCAG GAACAATATG AACTGGTCTA 900
CAATGCTGTA TTAGAACTAT TTAAGAGACA GATGGATGTT ATCAGAGATA AACATTCTGG 960
AACAGAGAGT CAAGCAAAGC ATTGTATTCC TGAGAAAAAT CACACTCTCC AAGCAGACTC 1020
TTATTCTCCT AATTTACCAA AAAGTACCAC AAAAGCAGCA AAAATGATGA ACCAACAAAG 1080
GACAAAAATG GAAATCAAAG AATCTTCTTC CTTTGACTTT AGGACTTCTG AAATAAGTGC 1140
AAAAGAAGAG CTAGTTTTGC ACCCTGCTAA ATCAAGCACT TCTTTTGACT TTCTGGAGCT 1200
AAATTACAGT TTTGACAAAA ATGCTGACAC AACCATGAAA TGGCAGACAA AGGCATTTCC 1260
AATAGTTGGG GAGCCTCTTC AGAAGCATCA AAGTTTGGAT TTGGGCTCTC TTTTGTTTGA 1320
GGGATGTTCT AATTCTAAAC CTGTAAATGC AGCAGGAAGA TATTTTAATT CAAAGGTGCC 1380
AATAACACGG ACCAAATCAA CTCCTTTTGA ATTGATACAG CAGAGAGAAA CCAAGGAGGT 1440
GGACAGCAAG GAAAACTTTT CTTATTTGGA ATCTCAACCA CATGATTCTT GTTTTGTAGA 1500
GATGCAGGCT CAAAAAGTAA TGCATGTTTC TTCAGCAGAA CTGAATTATT CACTGCCATA 1560
TGACTCTAAA CACCAAATAC GTAATGCCTC TAATGTAAAG CACCATGACT CTAGTGCTCT 1620
TGGTGTATAT TCTTACATAC CTTTAGTGGA AAATCCTTAT TTTTCATCAT GGCCTCCAAG 1680
TGGTACCAGT TCTAAGATGT CTCTTGATTT ACCTGAGAAG CAAGATGGAA CTGTTTTTCC 1740
TTCTTCTCTG TTGCCAACAT CCTCTACATC CCTCTTCTCT TATTACAATT CACATGATTC 1800
TTTATCACTG AATTCTCCAA CCAATATTTC CTCACTATTG AACCAGGAGT CAGCTGTACT 1860
AGCAACTGCT CCAAGGATAG ATGATGAAAT CCCCCCTCCA CTTCCTGTAC GGACACCTGA 1920
ATCATTTATT GTGGTTGAGG AAGCTGGAGA ATTCTCACCA AATGTTCCCA AATCCTTATC 1980
CTCAGCTGTG AAGGTAAAAA TTGGAACATC ACTGGAATGG GGTGGAACAT CTGAACCAAA 2040
GAAATTTGAT GACTCTGTGA TACTTAGACC AAGCAAGAGT GTAAAACTCC GAAGTCCTAA 2100
ATCAGAACTA CATCAAGATC GTTCTTCTCC CCCACCTCCT CTCCCAGAAA GAACTCTAGA 2160
GTCCTTCTTT CTTGCCGATG AAGATTGTAT GCAGGCCCAA TCTATAGAAA CATATTCTAC 2220
TAGCTATCCT GACACCATGG AAAATTCAAC ATCTTCAAAA CAGACACTGA AGACTCCTGG 2280

AAAAAGTTTC ACAAGGAGTA AGAGTTTGAA AATTTTGCGA AACATGAAAA AGAGTATCTG 2340
TAATTCTTGC CCACCAAACA AGCCTGCAGA ATCTGTTCAG TCAAATAACT CCAGCTCATT 2400
TCTGAATTTT GGTTTTGCAA ACCGTTTTTC AAAACCCAAA GGACCAAGGA ATCCACCACC 2460
AACTTGGAAT ATTTAATAAA ACTCCAGATT TATAATAATA TGGGCTGCAA GTACACCTGC 2520
AAATAAAACT ACTAGAATAC TGCTAGTTAA AATAAGTGCT CTATATGCAT AATATCAAAT 2580
ATGAAGATAT GCTAATGTGT TAATAGCTTT TAAAAGAAAA GCAAAATGCC AATAAGTGCC 2640
AGTTTTGCAT TTTCATATCA TTTGCATTGA GTTGAAAACT GCAAATAAAA GTTTGTCACT 2700
TGAGCTTATG TACAGAATGC TATATGAGAA ACACTTTTAG AATGGATTTA TTTTTCATTT 2760
TTGCCAGTTA TTTTTATTTT CTTTTACTTT TTTACATAAA CATAAACTTC AAAAGGTTTG 2820
TAAGATTTGG ATCTCAACTA ATTTCTACAT TGCCAGAATA TACTATAAAA AGTTAAAAAA 2880
AAACTTACTT TGTGGGTTGC AATACAAACT GCTCTTGACA ATGACTATTC CCTGACAGTT 2940
ATTTTTGCCT AAATGGAGTA TACCTTGTAA ATCTTCCCAA ATGTTGTGGA AAACTGGAAT 3000
ATTAAGAAAA TGAGAAATTA TATTTATTAG AATAAAATGT GCAAATAATG ACAATTATTT 3060
GAATGTAACA AGGAATTCAA CTGAAATCCT GATAAGTTTT AACCAAAGTC ATTAAATTAC 3120
CAATTCTAGA AAAGTAATCA ATGAAATATA ATAGCTATCT TTTGGTAGCA AAAGATATAA 3180
ATTGTATATG TTTATACAGG ATCTTTCAGA TCATGTGCAA TTTTTATCTA ACCAATCAGA 3240
AATACTAGTT TAAAATGAAT TTCTATATGA ATATGGATCT GCCATAAGAA AATCTAGTTC 3300
AACTCTAATT TTATGTAGTA AATAAATTGG CAGGTAATTG TTTTTACAAA GAATCCACCT 3360
GACTTCCCCT AATGCATTAA AAATATTTTT ATTTAAATAA CTTTATTTAT AACTTTTAGA 3420
AACATGTAGT ATTGTTTAAA CATCATTTGT TCTTCAGTAT TTTTCATTTG GAAGTCCAAT 3480
AGGGCAAATT GAATGAAGTA TTATTATCTG TCTCTTGTAG TACAATGTAT CCAACAGACA 3540
CTCAATAAAC TTTTTGGTTG TTAAAAAAAA AAAAAAAAAA 3580

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1548 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GCTCGCGGGC TCCCATGGCC CTCGGGCCCA GCGTGGTGAC CCCGGGGGAT GGAGCCGTTC 60
CTCAGGAGGC GGCTGGCCTT CCTGTCCTTC TTCTGGGACA AGATCTGGCC GGCGGGCGGC 120
GAGCCGGACC ATGGCACCCC CGGGTCCCTG GACCCCAACA CTGACCCAGT GCCCACGCTC 180
CCCGCCGAGC CTTGCAGCCC CTTCCCTCAG CTCTTCCTTG CGCTCTATGA CTTCACGGCG 240
CGGTGTGGCG GGGAGCTGAG TGTCCGCCGC GGGGACAGGC TCTGTGCCCT CGAAGAGGGG 300
GGCGGCTACA TCTTCGCACG CAGGCTTTCG GGCCAGCCCA GCGCCGGGCT CGTGCCCATC 360
ACCCACGTGG CCAAGGCTTC TCCTGAGACG CTCTCAGACC AACCCTGGTA CTTTAGCGGG 420
GTCAGTCGGA CCCAGGCACA GCAGCTGCTC CTCTCCCCAC CCAACGAACC AGGGGCCTTC 480
CTCATCCGGC CCAGCGAGAG CAGCCTCGGG GGCTACTCAC TGTCAGTCCG GGCCCAGGCC 540
AAGGTCTGCC ACTACCGGGT CTCCATGGCA GCTGATGGCA GCCTCTACCT GCAGAAGGGA 600
CGGCTCTTTC CCGGCCTGGA GGAGCTGCTC ACCTACTACA AGGCCAACTG GAAGCTGATC 660
CAGAACCCCC TGCTGCAGCC CTGCATGCCC CAGAAGGCCC CGAGGCAGGA CGTGTGGGAG 720
CGGCCACACT CCGAATTCGC CCTTGGGAGG AAGCTGGGTG AAGGCTACTT TGGGGAGGTG 780
TGGGAAGGCC TGTGGCTGGG CTCCCTGCCC GTGGCGATCA AGGTCATCAA GTCAGCCAAC 840
ATGAAGCTCA CTGACCTCGC CAAGGAGATC CAGACACTGA AGGGCCTGCG GCACGAGCGG 900
CTCATCCGGC TGCACGCAGT GTGCTCGGGC GGGGAGCCTG TGTACATAGT CACGGAACTC 960
ATGCGCAAGG GGAACCTGCA GGCCTTCCTG GGCACCCCCG AGGGCCGGGC CCTGCGTCTG 1020
CCGCCACTCC TGGGCTTTGC CTGCCAGGTG GCTGAGGGCA TGAGCTACCT GGAGGAGCAG 1080
CGCGTTGTGC ACCGGGACTT GGCCGCCCGG AACGTGCTCG TGGACGACGG CCTGGCCTGC 1140
AAGGTGGCTG ACTTCGGCCT GGCCCGGCTG CTCAAGGACG ACATCTACTC CCCGAGCAGC 1200
AGCTCCAAGA TCCCGGTCAA GTGGACAGCG CCTGAGGCGG CCAATTATCG TGTCTTCTCC 1260
CAGAAGTCAG ACGTCTGGTC CTTCGGCGTC CTGCTGCACG AGGTTTTCAC CTATGGCCAG 1320
TGTCCCTATG AAGGGATGAC CAACCACGAG ACGCTGCAGC AGATCATGCG AGGGTACCGG 1380
CTGCCGCGCC CGGCTGCCTG CCCGGCGGAG GTCTACGTGC TCATGCTGGA GTGCTGGAGG 1440
AGCAGCCCCG AGGAACGGCC CTCCTTTGCC ACGCTGCGGG AGAAGCTGCA CGCCATCCAC 1500
AGATGCCACC CCTGAGTCCT CACGTGACCC AACGCTCTGG GCTCCAGC 1548

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1785 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGTTATGTCT GACTCACTGC ACTGGAGTTT GGCAAAAGCA TCTCAGAAGT GGTTGTGCTT 60
TTTTGAATGA AATGATCAAT GGAGTGCTCC AGTTGTATGC TGGCCTCTGG ATACTAACTA 120
GACCTGCCTG ACTCCAGGAA CTAAGGCTCA GTATCTGCAG AAGCTTTTTG CCCATCTCAT 180
TCCGGCTATG GGGACAACAT GTCTTCACCC AGGAAGGTTA GAGGAAAAAC TGGAAGAGAT 240
AATGATGAAG AGGAGGGTAA TTCAGGTAAC CTGAATCTCC GCAACTCTTT GCCTTCATCG 300
AGTCAGAAAA TGACGCCTAC GAAGCCGATT TTTGGGAATA AAATGAATTC AGAGAATGTA 360
AAACCCTCCC ATCACCTGTC ATTCTCAGAT AAGTATGAGC TTGTTTACCC AGAGCCTTTG 420
GAAAGTGACA CTGATGAGAC TGTGTGGGAT GTCAGTGACC GGTCTCTCAG AAACAGGTGG 480
AACAGTATGG ATTCAGAGAC TGCAGGGCCG TCAAAGACTG TCTCCCCAGT GCTTTCTGGT 540
AGTAGTAGGC TCTCAAAGGA CACTGAAACA TCTGTCTCTG AAAAGGAGCT AACTCAGTTG 600
GCTCAGATTC GACCATTAAT ATTCAACAGT TCTGCACGGT CTGCTATGCG GGATTGTTTG 660
AACACGCTTC AGAAAAAAGA AGAACTTGAT ATCATCCGTG AGTTTTTGGA GTTAGAACAA 720
ATGACTCTGC CTGATGACTT CAATTCTGGG AATACACTAC AGAACAGAGA TAAGAACAGA 780
TACCGAGATA TTCTTCCATA TGATTCAACA CGTGTTCCTC TTGGAAAAAA CAAGGACTAC 840
ATCAACGCTA GTTATATTAG AATAGTAAAT CATGAAGAAG AGTATTTTTA TATTGCCACT 900
CAAGGACCAT TGCCAGAAAC TATAGAAGAC TTTTGGCAAA TGGTTCTGGA AAATAATTGT 960
AATGTTATTG CTATGATAAC CAGAGAGATA GAATGTGGAG TTATCAAGTG TTACAGTTAC 1020
TGGCCCATTT CTCTGAAGGA GCCTTTGGAA TTCGAACACT TTAGTGTCTT TCTGGAGACC 1080
TTTCATGTAA CTCAATATTT CACCGTTCGA GTATTTCAGA TTGTGAAGAA GTCCACAGGA 1140
AAGAGCCAAT GTGTAAAACA CTTGCAGTTC ACCAAGTGGC CAGACCATGG CACTCCTGCC 1200
TCAGCAGATT TTTTCATAAA ATATGTCCGT TATGTGAGGA AGAGCCACAT TACAGGACCC 1260
CTCCTTGTTC ACTGCAGTGC TGGTGTAGGC CGAACAGGGG TGTTCATATG TGTGGATGTT 1320
GTGTTCTCTG CCATCGAGAA GAACTACTCT TTTGACATTA TGAACATAGT GACCCAGATG 1380
AGAAAGCAGC GCTGTGGCAT GATTCAAACC AAGGAGCAGT ACCAGTTTTG TTATGAAATT 1440
GTGCTTGAAG TTCTTCAGAA CCTTCTGGCT TTGTATTAAG AGAGACTTCT GCGCCTGTCC 1500
CTCGAGGTTA CCGAGCAGCT TGGAGCCTGA GCCGTGCTGA AGCGTCTGCG GGCCGTGCAG 1560
TCTGCCTTCT GATTTTTCTC TCTGAAAGTC CCTGAAGGTA GCACTACTGG GCACAGAGTG 1620
AACTGTTTCC ACTTGATCTT TCTGAACAAG AGCAAAATAC CCTCCATGCC TTCTACGGAA 1680
ACGGAAGTTG CATGAAACAA CCTCCGCTTG GCTGTCTGGT TTGTGGTATT ACAGAGCTTA 1740
ATAAAAGACT TAGATGTGAA AAAAAAAAAA AAAAAAAAAA AAAAA 1785

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1896 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGTTATGTCT GACTCACTGC ACTGGAGTTT GGCAAAAGCA TCTCAGAAGT GGTTGTGCTT 60
TTTTGAATGA AATGATCAAT GGAGTGCTCC AGTTGTATGC TGGCCTCTGG ATACTAACTA 120
GACCTGCCTG ACTCCAGGAA CTAAGGCTCA GTATCTGCAG AAGCTTTTTG CCCATCTCAT 180
TCCGGCTATG GGGACAACAT GTCTTCACCC AGGAAGGTTA GAGGAAAAAC TGGAAGAGAT 240
AATGATGAAG AGGAGGGTAA TTCAGGTAAC CTGAATCTCC GCAACTCTTT GCCTTCATCG 300
AGTCAGAAAA TGACGCCTAC GAAGCCGGTA CAAAATAAAA ATCTCATGAA GTATGAAGAA 360
CACTTAGATA TATTGATGGT GTTTTTATTG ATAAAAACCA TATGGTATAA TGTCTTCAAA 420
TTATGGAAAG GCAAGCTTAT TTTTGGGAAT AAAATGAATT CAGAGAATGT AAAACCCTCC 480

```
CATCACCTGT CATTCTCAGA TAAGTATGAG CTTGTTTACC CAGAGCCTTT GGAAAGTGAC   540
ACTGATGAGA CTGTGTGGGA TGTCAGTGAC CGGTCTCTCA GAAACAGGTG GAACAGTATG   600
GATTCAGAGA CTGCAGGGCC GTCAAAGACT GTCTCCCCAG TGCTTTCTGG TAGTAGTAGG   660
CTCTCAAAGG ACACTGAAAC ATCTGTCTCT GAAAAGGAGC TAACTCAGTT GGCTCAGATT   720
CGACCATTAA TATTCAACAG TTCTGCACGG TCTGCTATGC GGGATTGTTT GAACACGCTT   780
CAGAAAAAAG AAGAACTTGA TATCATCCGT GAGTTTTTGG AGTTAGAACA AATGACTCTG   840
CCTGATGACT TCAATTCTGG GAATACACTA CAGAACAGAG ATAAGAACAG ATACCGAGAT   900
ATTCTTCCAT ATGATTCAAC ACGTGTTCCT CTTGGAAAAA ACAAGGACTA CATCAACGCT   960
AGTTATATTA GAATAGTAAA TCATGAAGAA GAGTATTTTT ATATTGCCAC TCAAGGACCA  1020
TTGCCAGAAA CTATAGAAGA CTTTTGGCAA ATGGTTCTGG AAAATAATTG TAATGTTATT  1080
GCTATGATAA CCAGAGAGAT AGAATGTGGA GTTATCAAGT GTTACAGTTA CTGGCCCATT  1140
TCTCTGAAGG AGCCTTTGGA ATTCGAACAC TTTAGTGTCT TTCTGGAGAC CTTTCATGTA  1200
ACTCAATATT TCACCGTTCG AGTATTTCAG ATTGTGAAGA AGTCCACAGG AAAGAGCCAA  1260
TGTGTAAAAC ACTTGCAGTT CACCAAGTGG CCAGACCATG GCACTCCTGC CTCAGCAGAT  1320
TTTTTCATAA AATATGTCCG TTATGTGAGG AAGAGCCACA TTACAGGACC CCTCCTTGTT  1380
CACTGCAGTG CTGGTGTAGG CCGAACAGGG GTGTTCATAT GTGTGGATGT TGTGTTCTCT  1440
GCCATCGAGA AGAACTACTC TTTTGACATT ATGAACATAG TGACCCAGAT GAGAAAGCAG  1500
CGCTGTGGCA TGATTCAAAC CAAGGAGCAG TACCAGTTTT GTTATGAAAT TGTGCTTGAA  1560
GTTCTTCAGA ACCTTCTGGC TTTGTATTAA GAGAGACTTC TGCGCCTGTC CCTCGAGGTT  1620
ACCGAGCAGC TTGGAGCCTG AGCCGTGCTG AAGCGTCTGC GGGCCGTGCA GTCTGCCTTC  1680
TGATTTTTCT CTCTGAAAGT CCCTGAAGGT AGCACTACTG GGCACAGAGT GAACTGTTTC  1740
CACTTGATCT TTCTGAACAA GAGCAAAATA CCCTCCATGC CTTCTACGGA AACGGAAGTT  1800
GCATGAAACA ACCTCCGCTT GGCTGTCTGG TTTGTGGTAT TACAGAGCTT AATAAAAGAC  1860
TTAGATGTGA AAAAAAAAAA AAAAAAAAAA AAAAAA                            1896
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	1692 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
GGTTATGTCT GACTCACTGC ACTGGAGTTT GGCAAAAGCA TCTCAGAAGT GGTTGTGCTT     60
TTTTGAATGA AATGATCAAT GGAGTGCTCC AGTTGTATGC TGGCCTCTGG ATACTAACTA    120
GACCTGCCTG ACTCCAGGAA CTAAGGCTCA GTATCTGCAG AAGCTTTTTG CCCATCTCAT    180
TCCGGCTATG GGGACAACAT GTCTTCACCC AGGAAGGTTA GAGGAAAAAC TGGAAGAGAT    240
AATGATGAAG AGGAGGGTAA TTCAGGTAAC CTGAATCTCC GCAACTCTTT GCCTTCATCG    300
AGTCAGAAAA TGACGCCTAC GAAGCCGATT TTTGGGAATA AAATGAATTC AGAGAATGTA    360
AAACCCTCCC ATCACCTGTC ATTCTCAGAT AAGTATGAGC TTGTTTACCC AGAGCCTTTG    420
GAAAGTGACA CTGATGAGAC TGTGTGGGAT GTCAGTGACC GGTCTCTCAG AAACAGGTGG    480
AACAGTATGG ATTCAGAGAC TGCAGGGCCG TCAAAGACTG TCTCCCCAGT GCTTTCTGGT    540
AGTAGTAGGC TCTCAAAGGA CACTGAAACA TCTGTCTCTG AAAAGGAGCT AACTCAGTTG    600
GCTCAGATTC GACCATTAAT ATTCAACAGT TCTGCACGGT CTGCTATGCG GGATTGTTTG    660
AACACGCTTC AGAAAAAAGA AGAACTTGAT ATCATCCGTG AGTTTTTGGA GTTAGAACAA    720
ATGACTCTGC CTGATGACTT CAATTCTGGG AATACACTAC AGAACAGAGA TAAGAACAGA    780
TACCGAGATA TTCTTCCATA TGATTCAACA CGTGTTCCTC TTGGAAAAAA CAAGGACTAC    840
ATCAACGCTA GTTATATTAG AATAGTAAAT CATGAAGAAG AGTATTTTTA TATTGCCACT    900
CAAGGACCAT TGCCAGAAAC TATAGAAGAC TTTTGGCAAA TGGTTCTGGA AAATAATTGT    960
AATGTTATTG CTATGATAAC CAGAGAGATA GAATGTGGAG TTATCAAGTG TTACAGTTAC   1020
TGGCCCATTT CTCTGAAGGA GCCTTTGGAA TTCGAACACT TTAGTGTCTT TCTGGAGACC   1080
TTTCATGTAA CTCAATATTT CACCGTTCGA GTATTTCAGA TTGTGAAGAA GTCCACAGGA   1140
AAGAGCCAAT GTGTAAAACA CTTGCAGTTC ACCAAGTGGC CAGACCATGG CACTCCTGCC   1200
TCAGCAGATT TTTTCATAAA ATATGTCCGT TATGTGAGGA AGAGCCACAT TACAGGACCC   1260
CTCCTTGTTC ACTGCAGTGC TGGTGTAGGC CGAACAGGGG TGTTCATATG TGTGGATGTT   1320
GTGTTCTCTG CCATCGAGAA GAACTACTCT TTTGACATTA TGAACATAGT GACCCAGATG   1380
AGAAAGCAGC GCTGTGGCAT GATTCAAACC AAGGTTACCG AGCAGCTTGG AGCCTGAGCC   1440
GTGCTGAAGC GTCTGCGGGC CGTGCAGTCT GCCTTCTGAT TTTTCTCTCT GAAAGTCCCT   1500
GAAGGTAGCA CTACTGGGCA CAGAGTGAAC TGTTTCCACT TGATCTTTCT GAACAAGAGC   1560
AAAATACCCT CCATGCCTTC TACGGAAACG GAAGTTGCAT GAAACAACCT CCGCTTGGCT   1620
GTCTGGTTTG TGGTATTACA GAGCTTAATA AAAGACTTAG ATGTGAAAAA AAAAAAAAAA   1680
```

AAAAAAAAAA AA 1692

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 320 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GAAAATAATT GTAATGTTAT TGCTATGATA ACCAGAGAGA TAGAAGGTGG AGTTATCAAG 60
TGTTGCAGTT ACTGGCCCGT TTCTCTGAAG GAGCCTTTGG AATTCAAACA CTTTCATGTC 120
CTTCTGGAGA ACTTTCAGAT AACTCAGTAT TTTGTCATCC GAATATTTCA AATTGTGAAG 180
AAGTCCACAG GAAAGAGTCA CTCTGTAAAA CACTTGCAGT TCATCAAATG GCCAGACCAT 240
GGCACTCCTG CCTCAGTAGA TTTTTTCATC AAATATGTCC GTTATGTGAG GAAGAGCCAC 300
ATTACAGGAC CCCTCCTTGT 320

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4456 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGCACGAGAG GAGCAGCAGA AGTTCGGGGA GCGGGTTGCA TACTTCCAGA GCGCCCTGGA 60
CAAGCTCAAT GAAGCCATCA AGTTGGCCAA GGGCCAGCCT GACACTGTGC AAGACGCGCT 120
TCGCTTCACT ATGGATGTCA TTGGGGGAAA GTACAATTCT GCCAAGAAGG ACAACGACTT 180
CATTTACCAT GAGGCTGTCC CAGCATTGAC ACCCTTCAGC CTGTAAAAGG AGCCCCCTTG 240
GTGAAGCCCT TGCCAGTGAA CCCCACAGAC CCAGCTGTTA CAGGCCCTGA CATCTTTGCC 300
AAACTGGTAC CCATGGCTGC CCACGAGGCC TCGTCACTGT ACAGTGAGGA GAAGGCCAAG 360
CTGCTCCGGG AGATGATGGC CAAGATTGAG GACAAGAATG AGGTCCTGGA CCAGTTCATG 420
GATTCAATGC AGTTGGATCC CGAGACGGTG GACAACCTTG ATGCCTACAG CCACATCCCA 480
CCCCAGCTCA TGGAGAAGTG CGCGGCTCTC AGCGTCCGGC CCGACACTGT CAGGAACCTT 540
GTACAGTCCA TGCAAGTGCT GTCAGGTGTG TTCACGGATG TGGAGGCTTC CCTGAAGGAC 600
ATCAGAGATC TGTTGGAGGA GGATGAGCTG CTAGAGCAGA AGTTTCAGGA GGCGGTGGGC 660
CAGGCAGGGG CCATCTCCAT CACCTCCAAG GCTGAGCTGG CAGAGGTGAG GCGAGAATGG 720
GCCAAGTACA TGGAAGTCCA TGAGAAGGCC TCCTTCACCA ACAGTGAGCT GCACCGTGCC 780
ATGAACCTGC ACGTCGGCAA CCTGCGCCTG CTCAGCGGGC CGCTTGACCA GGTCCGGGCT 840
GCCCTGCCCA CACCGGCCCT CTCCCCAGAG GACAAGGCCG TGCTGCAAAA CCTAAAGCGC 900
ATCCTGGCTA AGGTGCAGGA GATGCGGGAC CAGCGCGTGT CCCTGGAGCA GCAGCTGCGT 960
GAGCTTATCC AGAAAGATGA CATCACTGCC TCGCTGGTCA CCACAGACCA CTCAGAGATG 1020
AAGAAGTTGT TCGAGGAGCA GCTGAAAAAG TATGACCAGC TGAAGGTGTA CCTGGAGCAG 1080
AACCTGGCCG CCCAGGACCG TGTCCTCTGT GCACTGACAG AGGCCAACGT GCAGTACGCA 1140
GCCGTGCGGC GGGTACTCAG CGACTTGGAC CAAAAGTGGA ACTCCACGCT GCAGACCCTG 1200
GTGGCCTCGT ATGAAGCCTA TGAGGACCTG ATGAAGAAGT CGCAGGAGGG CAGGGACTTC 1260
TACGCAGATC TGGAGAGCAA GGTGGCTGCT CTGCTGGAGC GCACGCAGTC CACCTGCCAG 1320
GCCCGCGAGG CTGCCCGCCA GCAGCTCCTG GACAGGGAGC TGAAGAAGAA GCCGCCGCCA 1380
CGGCCCACAG CCCCAAAGCC GCTGCTGCCC CGCAGGGAGG AGAGTGAGGC AGTGGAAGCA 1440
GGAGACCCCC CTGAGGAGCT GCGCAGCCTC CCCCCTGACA TGGTGGCTGG CCCACGACTG 1500
CCTGACACCT TCCTGGGAAG TGCCACCCCG CTCCACTTTC CTCCCAGCCC CTTCCCCAGC 1560
TCCACAGGCC CAGGACCCCA CTATCTCTCA GGCCCCTTGC CCCCTGGTAC CTACTCGGGC 1620
CCCACCCAGC TGATACAGCC CAGGGCCCCA GGGCCCCATG CAATGCCCGT AGCACCTGGG 1680
CCTGCCCTCT ACCCAGCCCC TGCCTACACA CCGGAGCTGG GCCTTGTGCC CCGATCCTCC 1740
CCACAGCATG GCGTGGTGAG CAGTCCCTAT GTGGGGGTAG GGCCGGCCCC ACCAGTTGCA 1800
GGTCTCCCCT CGGCCCCACC TCCTCAATTC TCAGGCCCCG AGTTGGCCAT GGCGGTTCGG 1860
CCAGCCACCA CCACAGTAGA TAGCATCCAG GCGCCCATCC CCAGCCACAC AGCCCCACGG 1920

CCAAACCCCA CCCCTGCTCC TCCCCCGCCC TGCTTCCCTG TGCCCCCACC GCAGCCACTG 1980
CCCACGCCTT ACACCTACCC TGCAGGGGCT AAGCAACCCA TCCCAGCACA GCACCACTTC 2040
TCTTCTGGGA TCCCCACAGG TTTTCCAGCC CCAAGGATTG GGCCCCAGCC CCAGCCCCAT 2100
CCTCAGCCCC ATCCTTCACA AGCGTTTGGG CCTCAGCCCC CACAGCAGCC CCTTCCACTC 2160
CAGCATCCAC ATCTCTTCCC ACCCCAGGCC CCAGGACTCC TACCCCCACA ATCCCCCTAC 2220
CCCTATGCCC CTCAGCCTGG GGTCCTGGGG CAGCCGCCAC CCCCCCTACA CACCCAGCTC 2280
TACCCAGGTC CCGCTCAAGA CCCTCTGCCA GCCCACTCAG GGGCTCTGCC TTTCCCCAGC 2340
CCTGGGCCCC CTCAGCCTCC CCATCCCCCA CTGGCATATG GTCCTGCCCC TTCTACCAGA 2400
CCCATGGGCC CCCAGGCAGC CCCTCTTACC ATTCGAGGGC CCTCGTCTGC TGGCCAGTCC 2460
ACCCCTAGTC CCCACCTGGT GCCTTCACCT GCCCCATCTC CAGGGCCTGG TCCGGTACCC 2520
CCTCGCCCCC CAGCAGCAGA ACCACCCCCT TGCCTGCGCC GAGGCGCCGC AGCTGCAGAC 2580
CTGCTCTCCT CCAGCCCGGA GAGCCAGCAT GGCGGCACTC AGTCTCCTGG GGGTGGGCAG 2640
CCCCTGCTGC AGCCCACCAA GGTGGATGCA GCTGAGGGTC GTCGGCCGCA GGCCCTGCGG 2700
CTGATTGAGC GGGACCCCTA TGAGCATCCT GAGAGGCTGC GGCAGTTGCA GCAGGAGCTG 2760
GAGGCCTTTC GGGGTCAGCT GGGGGATGTG GGAGCTCTGG ACACTGTCTG GCGAGAGCTG 2820
CAAGATGCGC AGGAACATGA TGCCCGAGGC CGTTCCATCG CCATTGCCCG CTGCTACTCA 2880
CTGAAGAACC GGCACCAGGA TGTCATGCCC TATGACAGTA ACCGTGTGGT GCTGCGCTCA 2940
GGCAAGGATG ACTACATCAA TGCCAGCTGC GTGGAGGGGC TCTCCCCATA CTGCCCCCCG 3000
CTAGTGGCAA CCCAGGCCCC ACTGCCTGGC ACAGCTGCTG ACTTCTGGCT CATGGTCCAT 3060
GAGCAGAAAG TGTCAGTCAT TGTCATGCTG GTTTCTGAGG CTGAGATGGA GAAGCAAAAA 3120
GTGGCACGCT ACTTCCCCAC CGAGAGGGGC CAGCCCATGG TGCACGGTGC CCTGAGCCTG 3180
GCATTGAGCA GCGTCCGCAG CACCGAAACC CATGTGGAGC GCGTGCTGAG CCTGCAGTTC 3240
CGAGACCAGA GCCTCAAGCG CTCTCTTGTG CACCTGCACT TCCCCACTTG GCCTGAGTTA 3300
GGCCTGCCCG ACAGCCCCAG CAACTTGCTG CGCTTCATCC AGGAGGTGCA CGCACATTAC 3360
CTGCATCAGC GGCCGCTGCA CACGCCCATC ATTGTGCACT GCAGCTCTGG TGTGGGCCGC 3420
ACGGGAGCCT TTGCACTGCT CTATGCAGCT GTGCAGGAGG TGGAGGCTGG GAACGGAATC 3480
CCTGAGCTGC CTCAGCTGGT GCGGCGCATG CGGCAGCAGA GAAAGCACAT GCTGCAGGAG 3540
AAGCTGCACC TCAGGTTCTG CTATGAGGCA GTGGTGAGAC ACGTGGAGCA GGTCCTGCAG 3600
CGCCATGGTG TGCCTCCTCC ATGCAAACCC TTGGCCAGTG CAAGCATCAG CCAGAAGAAC 3660
CACCTTCCTC AGGACTCCCA GGACCTGGTC CTCGGTGGGG ATGTGCCCAT CAGCTCCATC 3720
CAGGCCACCA TTGCCAAGCT CAGCATTCGG CCTCCTGGGG GGTTGGAGTC CCCGGTTGCC 3780
AGCTTGCCAG GCCCTGCAGA GCCCCCAGGC CTCCCGCCAG CCAGCCTCCC AGAGTCTACC 3840
CCAATCCCAT CTTCCTCCCC ACCCCCCCTT TCCTCCCCAC TACCTGAGGC TCCCCAGCCT 3900
AAGGAGGAGC CGCCAGTGCC TGAAGCCCCC AGCTCGGGGC CCCCCTCCTC CTCCCTGGAA 3960
TTGCTGGCCT CCTTGACCCC AGAGGCCTTC TCCCTGGACA GCTCCCTGCG GGGCAAACAG 4020
CGGATGAGCA AGCATAACTT TCTGCAGGCC CATAACGGGC AAGGGCTGCG GGCCACCCGG 4080
CCCTCTGACG ACCCCCTCAG CCTTCTGGAT CCACTCTGGA CACTCAACAA GACCTGAACA 4140
GGTTTTGCCT ACCTGGTCCT TACACTACAT CATCATCATC TCATGCCCAC CTGCCCACAC 4200
CCAGCAGAGC TTCTCAGTGG GCACAGTCTC TTACTCCCAT TTCTGCTGCC TTTGGCCCTG 4260
CCTGGCCCAG CCTGCACCCC TGTGGGGTGG AAATGTACTG CAGGCTCTGG GTCAGGTTCT 4320
GCTCCTTTAT GGGACCCGAC ATTTTTCAGC TCTTTGCTAT TGAAATAATA AACCACCCTG 4380
TTCTGTGAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 4440
AAAAAAAAAA AAAAAA 4456

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1793 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CGGCCACACT GACTAGAGCC AACCGCGCAC TTCAAAAGGG TGTCGGTGCC GCGCTCCCCT 60
CCCGCGGCCC GGGAACTTCA AAGCGGGCCG TGCTGCCCCG GCTGCCTCGC TCTGCTCTGG 120
GGCCTCGCAG CCCCGGCGCG GCCGCCTGGT GGCGATGACC CGGGCGCTCT GCTCAGCGCT 180
CCGCCAGGCT CTCCTGCTGC TCGCAGCGGC CGCCGAGCTC TCGCCAGGAC TGAAGTGTGT 240
ATGTCTTTTG TGTGATTCTT CAAACTTTAC CTGCCAAACA GAAGGAGCAT GTTGGGCATC 300
AGTCATGCTA ACCAATGGAA AAGAGCAGGT GATCAAATCC TGTGTCTCCC TTCCAGAACT 360
GAATGCTCAA GTCTTCTGTC ATAGTTCCAA CAATGTTACC AAAACCGAAT GCTGCTTCAC 420
AGATTTTTGC AACAACATAA CACTGCACCT TCCAACAGCA TCACCAAATG CCCCAAAACT 480

```
TGGACCCATG GAGCTGGCCA TCATTATTAC TGTGCCTGTT TGCCTCCTGT CCATAGCTGC    540
GATGCTGACA GTATGGGCAT GCCAGGGTCG ACAGTGCTCC TACAGGAAGA AAAAGAGACC    600
AAATGTGGAG GAACCACTCT CTGAGTGCAA TCTGGTAAAT GCTGGAAAAA CTCTGAAAGA    660
TCTGATTTAT GATGTGACCG CCTCTGGATC TGGCTCTGGT CTACCTCTGT TGGTTCAAAG    720
GACAATTGCA AGGACGATTG TGCTTCAGGA AATAGTAGGA AAAGGTAGAT TTGGTGAGGT    780
GTGGCATGGA AGATGGTGTG GGGAAGATGT GGCTGTGAAA ATATTCTCCT CCAGAGATGA    840
AAGATCTTGG TTTCGTGAGG CAGAAATTTA CCAGACGGTC ATGCTGCGAC ATGAAAACAT    900
CCTTGGTTTC ATTGCTGCTG ACAACAAAGA TAATGGAACT TGGACTCAAC TTTGGCTGGT    960
ATCTGAATAT CATGAACAGG GCTCCTTATA TGACTATTTG AATAGAAATA TAGTGACCGT   1020
GGCTGGAATG ATCAAGCTGG CGCTCTCAAT TGCTAGTGGT CTGGCACACC TTCATATGGA   1080
GATTGTTGGT ACACAAGGTA AACCTGCTAT TGCTCATCGA GACATAAAAT CAAAGAATAT   1140
CTTAGTGAAA AAGTGTGAAA CTTGTGCCAT AGCGGACTTA GGGTTGGCTG TGAAGCATGA   1200
TTCAATACTG AACACTATCG ACATACCTCA GAATCCTAAA GTGGGAACCA AGAGGTATAT   1260
GGCTCCTGAA ATGCTTGATG ATACAATGAA TGTGAATATC TTTGAGTCCT TCAAACGAGC   1320
TGACATCTAT TCTGTTGGTC TGGTTTACTG GGAAATAGCC CGGAGGTGTT CAGTCGGAGG   1380
AATTGTTGAG GAGTACCAAT TGCCTTATTA TGACATGGTG CCTTCAGATC CCTCGATAGA   1440
GGAAATGAGA AAGGTTGTTT GTGACCAGAA GTTTCGACCA AGTATCCCAA ACCAGTGGCA   1500
AAGTTGTGAA GCACTCCGAG TCATGGGGAG AATAATGCGT GAGTGTTGGT ATGCCAACGG   1560
AGCGGCCCGC CTAACTGCTC TTCGTATTAA GAAGACTATA TCTCAACTTT GTGTCAAAGA   1620
AGACTGCAAA GCCTAATGAT GATAATTATG TTAAAAAGAA ATCTCTCATA GCTTTCTTTT   1680
CCATTTTCCC CTTTATGTGA ATGTTTTTGC CATTTTTTTT TTGTTCTACC TCAAAGATAA   1740
GACAGTACAG TATTTAAGTG CCCATAAGGC AGCATGAAAA GATAACTCTA AAG           1793
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	807 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
Met Asp Gln Arg Glu Ile Leu Gln Lys Phe Leu Asp Glu Ala Gln Ser
 1               5                  10                  15

Lys Lys Ile Thr Lys Glu Glu Phe Ala Asn Glu Phe Leu Lys Leu Lys
            20                  25                  30

Arg Gln Ser Thr Lys Tyr Lys Ala Asp Lys Thr Tyr Pro Thr Thr Val
        35                  40                  45

Ala Glu Lys Pro Lys Asn Ile Lys Lys Asn Arg Tyr Lys Asp Ile Leu
    50                  55                  60

Pro Tyr Asp Tyr Ser Arg Val Glu Leu Ser Leu Ile Thr Ser Asp Glu
65                  70                  75                  80

Asp Ser Ser Tyr Ile Asn Ala Asn Phe Ile Lys Gly Val Tyr Gly Pro
                85                  90                  95

Lys Ala Tyr Ile Ala Thr Gln Gly Pro Leu Ser Thr Thr Leu Leu Asp
            100                 105                 110

Phe Trp Arg Met Ile Trp Glu Tyr Ser Val Leu Ile Ile Val Met Ala
        115                 120                 125

Cys Met Glu Tyr Glu Met Gly Lys Lys Lys Cys Glu Arg Tyr Trp Ala
    130                 135                 140
```

```
Glu Pro Gly Glu Met Gln Leu Glu Phe Gly Pro Phe Ser Val Ser Cys
145                 150                 155                 160

Glu Ala Glu Lys Arg Lys Ser Asp Tyr Ile Ile Arg Thr Leu Lys Val
                165                 170                 175

Lys Phe Asn Ser Glu Thr Arg Thr Ile Tyr Gln Phe His Tyr Lys Asn
            180                 185                 190

Trp Pro Asp His Asp Val Pro Ser Ser Ile Asp Pro Ile Leu Glu Leu
        195                 200                 205

Ile Trp Asp Val Arg Cys Tyr Gln Glu Asp Asp Ser Val Pro Ile Cys
    210                 215                 220

Ile His Cys Ser Ala Gly Cys Gly Arg Thr Gly Val Ile Cys Ala Ile
225                 230                 235                 240

Asp Tyr Thr Trp Met Leu Leu Lys Asp Gly Ile Ile Pro Glu Asn Phe
                245                 250                 255

Ser Val Phe Ser Leu Ile Arg Glu Met Arg Thr Gln Arg Pro Ser Leu
            260                 265                 270

Val Gln Thr Gln Glu Gln Tyr Glu Leu Val Tyr Asn Ala Val Leu Glu
        275                 280                 285

Leu Phe Lys Arg Gln Met Asp Val Ile Arg Asp Lys His Ser Gly Thr
    290                 295                 300

Glu Ser Gln Ala Lys His Cys Ile Pro Glu Lys Asn His Thr Leu Gln
305                 310                 315                 320

Ala Asp Ser Tyr Ser Pro Asn Leu Pro Lys Ser Thr Thr Lys Ala Ala
                325                 330                 335

Lys Met Met Asn Gln Gln Arg Thr Lys Met Glu Ile Lys Glu Ser Ser
            340                 345                 350

Ser Phe Asp Phe Arg Thr Ser Glu Ile Ser Ala Lys Glu Glu Leu Val
        355                 360                 365

Leu His Pro Ala Lys Ser Ser Thr Ser Phe Asp Phe Leu Glu Leu Asn
    370                 375                 380

Tyr Ser Phe Asp Lys Asn Ala Asp Thr Thr Met Lys Trp Gln Thr Lys
385                 390                 395                 400

Ala Phe Pro Ile Val Gly Glu Pro Leu Gln Lys His Gln Ser Leu Asp
                405                 410                 415

Leu Gly Ser Leu Leu Phe Glu Gly Cys Ser Asn Ser Lys Pro Val Asn
            420                 425                 430

Ala Ala Gly Arg Tyr Phe Asn Ser Lys Val Pro Ile Thr Arg Thr Lys
        435                 440                 445

Ser Thr Pro Phe Glu Leu Ile Gln Gln Arg Glu Thr Lys Glu Val Asp
    450                 455                 460

Ser Lys Glu Asn Phe Ser Tyr Leu Glu Ser Gln Pro His Asp Ser Cys
465                 470                 475                 480

Phe Val Glu Met Gln Ala Gln Lys Val Met His Val Ser Ser Ala Glu
                485                 490                 495
```

Leu Asn Tyr Ser Leu Pro Tyr Asp Ser Lys His Gln Ile Arg Asn Ala
500 505 510

Ser Asn Val Lys His His Asp Ser Ser Ala Leu Gly Val Tyr Ser Tyr
515 520 525

Ile Pro Leu Val Glu Asn Pro Tyr Phe Ser Ser Trp Pro Pro Ser Gly
530 535 540

Thr Ser Ser Lys Met Ser Leu Asp Leu Pro Glu Lys Gln Asp Gly Thr
545 550 555 560

Val Phe Pro Ser Ser Leu Leu Pro Thr Ser Ser Thr Ser Leu Phe Ser
565 570 575

Tyr Tyr Asn Ser His Asp Ser Leu Ser Leu Asn Ser Pro Thr Asn Ile
580 585 590

Ser Ser Leu Leu Asn Gln Glu Ser Ala Val Leu Ala Thr Ala Pro Arg
595 600 605

Ile Asp Asp Glu Ile Pro Pro Pro Leu Pro Val Arg Thr Pro Glu Ser
610 615 620

Phe Ile Val Val Glu Glu Ala Gly Glu Phe Ser Pro Asn Val Pro Lys
625 630 635 640

Ser Leu Ser Ser Ala Val Lys Val Lys Ile Gly Thr Ser Leu Glu Trp
645 650 655

Gly Gly Thr Ser Glu Pro Lys Lys Phe Asp Asp Ser Val Ile Leu Arg
660 665 670

Pro Ser Lys Ser Val Lys Leu Arg Ser Pro Lys Ser Glu Leu His Gln
675 680 685

Asp Arg Ser Ser Pro Pro Pro Pro Leu Pro Glu Arg Thr Leu Glu Ser
690 695 700

Phe Phe Leu Ala Asp Glu Asp Cys Met Gln Ala Gln Ser Ile Glu Thr
705 710 715 720

Tyr Ser Thr Ser Tyr Pro Asp Thr Met Glu Asn Ser Thr Ser Ser Lys
725 730 735

Gln Thr Leu Lys Thr Pro Gly Lys Ser Phe Thr Arg Ser Lys Ser Leu
740 745 750

Lys Ile Leu Arg Asn Met Lys Lys Ser Ile Cys Asn Ser Cys Pro Pro
755 760 765

Asn Lys Pro Ala Glu Ser Val Gln Ser Asn Asn Ser Ser Ser Phe Leu
770 775 780

Asn Phe Gly Phe Ala Asn Arg Phe Ser Lys Pro Lys Gly Pro Arg Asn
785 790 795 800

Pro Pro Pro Thr Trp Asn Ile
805

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 488 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Glu Pro Phe Leu Arg Arg Arg Leu Ala Phe Leu Ser Phe Phe Trp
1 5 10 15

Asp Lys Ile Trp Pro Ala Gly Gly Glu Pro Asp His Gly Thr Pro Gly
20 25 30

Ser Leu Asp Pro Asn Thr Asp Pro Val Pro Thr Leu Pro Ala Glu Pro
35 40 45

Cys Ser Pro Phe Pro Gln Leu Phe Leu Ala Leu Tyr Asp Phe Thr Ala
50 55 60

Arg Cys Gly Gly Glu Leu Ser Val Arg Arg Gly Asp Arg Leu Cys Ala
65 70 75 80

Leu Glu Glu Gly Gly Gly Tyr Ile Phe Ala Arg Arg Leu Ser Gly Gln
85 90 95

Pro Ser Ala Gly Leu Val Pro Ile Thr His Val Ala Lys Ala Ser Pro
100 105 110

Glu Thr Leu Ser Asp Gln Pro Trp Tyr Phe Ser Gly Val Ser Arg Thr
115 120 125

Gln Ala Gln Gln Leu Leu Leu Ser Pro Pro Asn Glu Pro Gly Ala Phe
130 135 140

Leu Ile Arg Pro Ser Glu Ser Ser Leu Gly Gly Tyr Ser Leu Ser Val
145 150 155 160

Arg Ala Gln Ala Lys Val Cys His Tyr Arg Val Ser Met Ala Ala Asp
165 170 175

Gly Ser Leu Tyr Leu Gln Lys Gly Arg Leu Phe Pro Gly Leu Glu Glu
180 185 190

Leu Leu Thr Tyr Tyr Lys Ala Asn Trp Lys Leu Ile Gln Asn Pro Leu
195 200 205

Leu Gln Pro Cys Met Pro Gln Lys Ala Pro Arg Gln Asp Val Trp Glu
210 215 220

Arg Pro His Ser Glu Phe Ala Leu Gly Arg Lys Leu Gly Glu Gly Tyr
225 230 235 240

Phe Gly Glu Val Trp Glu Gly Leu Trp Leu Gly Ser Leu Pro Val Ala
245 250 255

Ile Lys Val Ile Lys Ser Ala Asn Met Lys Leu Thr Asp Leu Ala Lys
260 265 270

```
Glu Ile Gln Thr Leu Lys Gly Leu Arg His Glu Arg Leu Ile Arg Leu
        275                 280                 285

His Ala Val Cys Ser Gly Gly Glu Pro Val Tyr Ile Val Thr Glu Leu
    290                 295                 300

Met Arg Lys Gly Asn Leu Gln Ala Phe Leu Gly Thr Pro Glu Gly Arg
305                 310                 315                 320

Ala Leu Arg Leu Pro Pro Leu Leu Gly Phe Ala Cys Gln Val Ala Glu
                325                 330                 335

Gly Met Ser Tyr Leu Glu Glu Gln Arg Val Val His Arg Asp Leu Ala
            340                 345                 350

Ala Arg Asn Val Leu Val Asp Asp Gly Leu Ala Cys Lys Val Ala Asp
        355                 360                 365

Phe Gly Leu Ala Arg Leu Leu Lys Asp Asp Ile Tyr Ser Pro Ser Ser
    370                 375                 380

Ser Ser Lys Ile Pro Val Lys Trp Thr Ala Pro Glu Ala Ala Asn Tyr
385                 390                 395                 400

Arg Val Phe Ser Gln Lys Ser Asp Val Trp Ser Phe Gly Val Leu Leu
                405                 410                 415

His Glu Val Phe Thr Tyr Gly Gln Cys Pro Tyr Glu Gly Met Thr Asn
            420                 425                 430

His Glu Thr Leu Gln Gln Ile Met Arg Gly Tyr Arg Leu Pro Arg Pro
        435                 440                 445

Ala Ala Cys Pro Ala Glu Val Tyr Val Leu Met Leu Glu Cys Trp Arg
    450                 455                 460

Ser Ser Pro Glu Glu Arg Pro Ser Phe Ala Thr Leu Arg Glu Lys Leu
465                 470                 475                 480

His Ala Ile His Arg Cys His Pro
                485
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
Met Ser Ser Pro Arg Lys Val Arg Gly Lys Thr Gly Arg Asp Asn Asp
 1               5                  10                  15

Glu Glu Glu Gly Asn Ser Gly Asn Leu Asn Leu Arg Asn Ser Leu Pro
            20                  25                  30

Ser Ser Ser Gln Lys Met Thr Pro Thr Lys Pro Ile Phe Gly Asn Lys
        35                  40                  45
```

```
Met Asn Ser Glu Asn Val Lys Pro Ser His His Leu Ser Phe Ser Asp
 50                  55                  60

Lys Tyr Glu Leu Val Tyr Pro Glu Pro Leu Glu Ser Asp Thr Asp Glu
65                  70                  75                  80

Thr Val Trp Asp Val Ser Asp Arg Ser Leu Arg Asn Arg Trp Asn Ser
                 85                  90                  95

Met Asp Ser Glu Thr Ala Gly Pro Ser Lys Thr Val Ser Pro Val Leu
            100                 105                 110

Ser Gly Ser Ser Arg Leu Ser Lys Asp Thr Glu Thr Ser Val Ser Glu
        115                 120                 125

Lys Glu Leu Thr Gln Leu Ala Gln Ile Arg Pro Leu Ile Phe Asn Ser
    130                 135                 140

Ser Ala Arg Ser Ala Met Arg Asp Cys Leu Asn Thr Leu Gln Lys Lys
145                 150                 155                 160

Glu Glu Leu Asp Ile Ile Arg Glu Phe Leu Glu Leu Glu Gln Met Thr
                165                 170                 175

Leu Pro Asp Asp Phe Asn Ser Gly Asn Thr Leu Gln Asn Arg Asp Lys
            180                 185                 190

Asn Arg Tyr Arg Asp Ile Leu Pro Tyr Asp Ser Thr Arg Val Pro Leu
        195                 200                 205

Gly Lys Asn Lys Asp Tyr Ile Asn Ala Ser Tyr Ile Arg Ile Val Asn
    210                 215                 220

His Glu Glu Glu Tyr Phe Tyr Ile Ala Thr Gln Gly Pro Leu Pro Glu
225                 230                 235                 240

Thr Ile Glu Asp Phe Trp Gln Met Val Leu Glu Asn Asn Cys Asn Val
                245                 250                 255

Ile Ala Met Ile Thr Arg Glu Ile Glu Cys Gly Val Ile Lys Cys Tyr
            260                 265                 270

Ser Tyr Trp Pro Ile Ser Leu Lys Glu Pro Leu Glu Phe Glu His Phe
        275                 280                 285

Ser Val Phe Leu Glu Thr Phe His Val Thr Gln Tyr Phe Thr Val Arg
    290                 295                 300

Val Phe Gln Ile Val Lys Lys Ser Thr Gly Lys Ser Gln Cys Val Lys
305                 310                 315                 320

His Leu Gln Phe Thr Lys Trp Pro Asp His Gly Thr Pro Ala Ser Ala
                325                 330                 335

Asp Phe Phe Ile Lys Tyr Val Arg Tyr Val Arg Lys Ser His Ile Thr
            340                 345                 350

Gly Pro Leu Leu Val His Cys Ser Ala Gly Val Gly Arg Thr Gly Val
        355                 360                 365

Phe Ile Cys Val Asp Val Val Phe Ser Ala Ile Glu Lys Asn Tyr Ser
    370                 375                 380
```

Phe Asp Ile Met Asn Ile Val Thr Gln Met Arg Lys Gln Arg Cys Gly
385 390 395 400

Met Ile Gln Thr Lys Glu Gln Tyr Gln Phe Cys Tyr Glu Ile Val Leu
405 410 415

Glu Val Leu Gln Asn Leu Leu Ala Leu Tyr
420 425

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 463 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Ser Ser Pro Arg Lys Val Arg Gly Lys Thr Gly Arg Asp Asn Asp
1 5 10 15

Glu Glu Glu Gly Asn Ser Gly Asn Leu Asn Leu Arg Asn Ser Leu Pro
20 25 30

Ser Ser Ser Gln Lys Met Thr Pro Thr Lys Pro Val Gln Asn Lys Asn
35 40 45

Leu Met Lys Tyr Glu Glu His Leu Asp Ile Leu Met Val Phe Leu Leu
50 55 60

Ile Lys Thr Ile Trp Tyr Asn Val Phe Lys Leu Trp Lys Gly Lys Leu
65 70 75 80

Ile Phe Gly Asn Lys Met Asn Ser Glu Asn Val Lys Pro Ser His His
85 90 95

Leu Ser Phe Ser Asp Lys Tyr Glu Leu Val Tyr Pro Glu Pro Leu Glu
100 105 110

Ser Asp Thr Asp Glu Thr Val Trp Asp Val Ser Asp Arg Ser Leu Arg
115 120 125

Asn Arg Trp Asn Ser Met Asp Ser Glu Thr Ala Gly Pro Ser Lys Thr
130 135 140

Val Ser Pro Val Leu Ser Gly Ser Ser Arg Leu Ser Lys Asp Thr Glu
145 150 155 160

Thr Ser Val Ser Glu Lys Glu Leu Thr Gln Leu Ala Gln Ile Arg Pro
165 170 175

Leu Ile Phe Asn Ser Ser Ala Arg Ser Ala Met Arg Asp Cys Leu Asn
180 185 190

Thr Leu Gln Lys Lys Glu Glu Leu Asp Ile Ile Arg Glu Phe Leu Glu
195 200 205

Leu Glu Gln Met Thr Leu Pro Asp Asp Phe Asn Ser Gly Asn Thr Leu
210 215 220

```
Gln Asn Arg Asp Lys Asn Arg Tyr Arg Asp Ile Leu Pro Tyr Asp Ser
225                 230                 235                 240

Thr Arg Val Pro Leu Gly Lys Asn Lys Asp Tyr Ile Asn Ala Ser Tyr
                245                 250                 255

Ile Arg Ile Val Asn His Glu Glu Glu Tyr Phe Tyr Ile Ala Thr Gln
            260                 265                 270

Gly Pro Leu Pro Glu Thr Ile Glu Asp Phe Trp Gln Met Val Leu Glu
        275                 280                 285

Asn Asn Cys Asn Val Ile Ala Met Ile Thr Arg Glu Ile Glu Cys Gly
    290                 295                 300

Val Ile Lys Cys Tyr Ser Tyr Trp Pro Ile Ser Leu Lys Glu Pro Leu
305                 310                 315                 320

Glu Phe Glu His Phe Ser Val Phe Leu Glu Thr Phe His Val Thr Gln
                325                 330                 335

Tyr Phe Thr Val Arg Val Phe Gln Ile Val Lys Lys Ser Thr Gly Lys
            340                 345                 350

Ser Gln Cys Val Lys His Leu Gln Phe Thr Lys Trp Pro Asp His Gly
        355                 360                 365

Thr Pro Ala Ser Ala Asp Phe Phe Ile Lys Tyr Val Arg Tyr Val Arg
    370                 375                 380

Lys Ser His Ile Thr Gly Pro Leu Leu Val His Cys Ser Ala Gly Val
385                 390                 395                 400

Gly Arg Thr Gly Val Phe Ile Cys Val Asp Val Val Phe Ser Ala Ile
                405                 410                 415

Glu Lys Asn Tyr Ser Phe Asp Ile Met Asn Ile Val Thr Gln Met Arg
            420                 425                 430

Lys Gln Arg Cys Gly Met Ile Gln Thr Lys Glu Gln Tyr Gln Phe Cys
        435                 440                 445

Tyr Glu Ile Val Leu Glu Val Leu Gln Asn Leu Leu Ala Leu Tyr
    450                 455                 460
```

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 405 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
Met Ser Ser Pro Arg Lys Val Arg Gly Lys Thr Gly Arg Asp Asn Asp
 1               5                  10                  15

Glu Glu Glu Gly Asn Ser Gly Asn Leu Asn Leu Arg Asn Ser Leu Pro
            20                  25                  30
```

Ser Ser Ser Gln Lys Met Thr Pro Thr Lys Pro Ile Phe Gly Asn Lys
35 40 45

Met Asn Ser Glu Asn Val Lys Pro Ser His His Leu Ser Phe Ser Asp
50 55 60

Lys Tyr Glu Leu Val Tyr Pro Glu Pro Leu Glu Ser Asp Thr Asp Glu
65 70 75 80

Thr Val Trp Asp Val Ser Asp Arg Ser Leu Arg Asn Arg Trp Asn Ser
85 90 95

Met Asp Ser Glu Thr Ala Gly Pro Ser Lys Thr Val Ser Pro Val Leu
100 105 110

Ser Gly Ser Ser Arg Leu Ser Lys Asp Thr Glu Thr Ser Val Ser Glu
115 120 125

Lys Glu Leu Thr Gln Leu Ala Gln Ile Arg Pro Leu Ile Phe Asn Ser
130 135 140

Ser Ala Arg Ser Ala Met Arg Asp Cys Leu Asn Thr Leu Gln Lys Lys
145 150 155 160

Glu Glu Leu Asp Ile Ile Arg Glu Phe Leu Glu Leu Glu Gln Met Thr
165 170 175

Leu Pro Asp Asp Phe Asn Ser Gly Asn Thr Leu Gln Asn Arg Asp Lys
180 185 190

Asn Arg Tyr Arg Asp Ile Leu Pro Tyr Asp Ser Thr Arg Val Pro Leu
195 200 205

Gly Lys Asn Lys Asp Tyr Ile Asn Ala Ser Tyr Ile Arg Ile Val Asn
210 215 220

His Glu Glu Glu Tyr Phe Tyr Ile Ala Thr Gln Gly Pro Leu Pro Glu
225 230 235 240

Thr Ile Glu Asp Phe Trp Gln Met Val Leu Glu Asn Asn Cys Asn Val
245 250 255

Ile Ala Met Ile Thr Arg Glu Ile Glu Cys Gly Val Ile Lys Cys Tyr
260 265 270

Ser Tyr Trp Pro Ile Ser Leu Lys Glu Pro Leu Glu Phe Glu His Phe
275 280 285

Ser Val Phe Leu Glu Thr Phe His Val Thr Gln Tyr Phe Thr Val Arg
290 295 300

Val Phe Gln Ile Val Lys Lys Ser Thr Gly Lys Ser Gln Cys Val Lys
305 310 315 320

His Leu Gln Phe Thr Lys Trp Pro Asp His Gly Thr Pro Ala Ser Ala
325 330 335

Asp Phe Phe Ile Lys Tyr Val Arg Tyr Val Arg Lys Ser His Ile Thr
340 345 350

Gly Pro Leu Leu Val His Cys Ser Ala Gly Val Gly Arg Thr Gly Val
355 360 365

Phe Ile Cys Val Asp Val Val Phe Ser Ala Ile Glu Lys Asn Tyr Ser
370 375 380

Phe Asp Ile Met Asn Ile Val Thr Gln Met Arg Lys Gln Arg Cys Gly
385 390 395 400

Met Ile Gln Thr Lys
405

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 122 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Asp Phe Trp Gly Met Met Trp Glu Asn Asn Cys Asn Val Ile Ala Met
1 5 10 15

Ile Thr Arg Glu Ile Glu Gly Gly Val Ile Lys Cys Cys Ser Tyr Trp
20 25 30

Pro Val Ser Leu Lys Glu Pro Leu Glu Phe Lys His Phe His Val Leu
35 40 45

Leu Glu Asn Phe Gln Ile Thr Gln Tyr Phe Val Ile Arg Ile Phe Gln
50 55 60

Ile Val Lys Lys Ser Thr Gly Lys Ser His Ser Val Lys His Leu Gln
65 70 75 80

Phe Ile Lys Trp Pro Asp His Gly Thr Pro Ala Ser Val Asp Phe Phe
85 90 95

Ile Lys Tyr Val Arg Tyr Val Arg Lys Ser His Ile Thr Gly Pro Leu
100 105 110

Leu Val His Cys Thr Ala Gly Val Gly Arg
115 120

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1274 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Met Ala Ala His Glu Ala Ser Ser Leu Tyr Ser Glu Glu Lys Ala Lys
1 5 10 15

Leu Leu Arg Glu Met Met Ala Lys Ile Glu Asp Lys Asn Glu Val Leu
20 25 30

Asp Gln Phe Met Asp Ser Met Gln Leu Asp Pro Glu Thr Val Asp Asn
35 40 45

Leu Asp Ala Tyr Ser His Ile Pro Pro Gln Leu Met Glu Lys Cys Ala
50 55 60

Ala Leu Ser Val Arg Pro Asp Thr Val Arg Asn Leu Val Gln Ser Met
65 70 75 80

Gln Val Leu Ser Gly Val Phe Thr Asp Val Glu Ala Ser Leu Lys Asp
85 90 95

Ile Arg Asp Leu Leu Glu Glu Asp Glu Leu Leu Glu Gln Lys Phe Gln
100 105 110

Glu Ala Val Gly Gln Ala Gly Ala Ile Ser Ile Thr Ser Lys Ala Glu
115 120 125

Leu Ala Glu Val Arg Arg Glu Trp Ala Lys Tyr Met Glu Val His Glu
130 135 140

Lys Ala Ser Phe Thr Asn Ser Glu Leu His Arg Ala Met Asn Leu His
145 150 155 160

Val Gly Asn Leu Arg Leu Leu Ser Gly Pro Leu Asp Gln Val Arg Ala
165 170 175

Ala Leu Pro Thr Pro Ala Leu Ser Pro Glu Asp Lys Ala Val Leu Gln
180 185 190

Asn Leu Lys Arg Ile Leu Ala Lys Val Gln Glu Met Arg Asp Gln Arg
195 200 205

Val Ser Leu Glu Gln Gln Leu Arg Glu Leu Ile Gln Lys Asp Asp Ile
210 215 220

Thr Ala Ser Leu Val Thr Thr Asp His Ser Glu Met Lys Lys Leu Phe
225 230 235 240

Glu Glu Gln Leu Lys Lys Tyr Asp Gln Leu Lys Val Tyr Leu Glu Gln
245 250 255

Asn Leu Ala Ala Gln Asp Arg Val Leu Cys Ala Leu Thr Glu Ala Asn
260 265 270

Val Gln Tyr Ala Ala Val Arg Arg Val Leu Ser Asp Leu Asp Gln Lys
275 280 285

Trp Asn Ser Thr Leu Gln Thr Leu Val Ala Ser Tyr Glu Ala Tyr Glu
290 295 300

Asp Leu Met Lys Lys Ser Gln Glu Gly Arg Asp Phe Tyr Ala Asp Leu
305 310 315 320

Glu Ser Lys Val Ala Ala Leu Leu Glu Arg Thr Gln Ser Thr Cys Gln
325 330 335

Ala Arg Glu Ala Ala Arg Gln Gln Leu Leu Asp Arg Glu Leu Lys Lys
340 345 350

Lys Pro Pro Pro Arg Pro Thr Ala Pro Lys Pro Leu Leu Pro Arg Arg
355 360 365

Glu Glu Ser Glu Ala Val Glu Ala Gly Asp Pro Pro Glu Glu Leu Arg
370 375 380

Ser Leu Pro Pro Asp Met Val Ala Gly Pro Arg Leu Pro Asp Thr Phe
385 390 395 400

Leu Gly Ser Ala Thr Pro Leu His Phe Pro Pro Ser Pro Phe Pro Ser
405 410 415

Ser Thr Gly Pro Gly Pro His Tyr Leu Ser Gly Pro Leu Pro Pro Gly
420 425 430

Thr Tyr Ser Gly Pro Thr Gln Leu Ile Gln Pro Arg Ala Pro Gly Pro
435 440 445

His Ala Met Pro Val Ala Pro Gly Pro Ala Leu Tyr Pro Ala Pro Ala
450 455 460

Tyr Thr Pro Glu Leu Gly Leu Val Pro Arg Ser Ser Pro Gln His Gly
465 470 475 480

Val Val Ser Ser Pro Tyr Val Gly Val Gly Pro Ala Pro Pro Val Ala
485 490 495

Gly Leu Pro Ser Ala Pro Pro Pro Gln Phe Ser Gly Pro Glu Leu Ala
500 505 510

Met Ala Val Arg Pro Ala Thr Thr Thr Val Asp Ser Ile Gln Ala Pro
515 520 525

Ile Pro Ser His Thr Ala Pro Arg Pro Asn Pro Thr Pro Ala Pro Pro
530 535 540

Pro Pro Cys Phe Pro Val Pro Pro Pro Gln Pro Leu Pro Thr Pro Tyr
545 550 555 560

Thr Tyr Pro Ala Gly Ala Lys Gln Pro Ile Pro Ala Gln His His Phe
565 570 575

Ser Ser Gly Ile Pro Thr Gly Phe Pro Ala Pro Arg Ile Gly Pro Gln
580 585 590

Pro Gln Pro His Pro Gln Pro His Pro Ser Gln Ala Phe Gly Pro Gln
595 600 605

Pro Pro Gln Gln Pro Leu Pro Leu Gln His Pro His Leu Phe Pro Pro
610 615 620

Gln Ala Pro Gly Leu Leu Pro Pro Gln Ser Pro Tyr Pro Tyr Ala Pro
625 630 635 640

Gln Pro Gly Val Leu Gly Gln Pro Pro Pro Pro Leu His Thr Gln Leu
645 650 655

Tyr Pro Gly Pro Ala Gln Asp Pro Leu Pro Ala His Ser Gly Ala Leu
660 665 670

Pro Phe Pro Ser Pro Gly Pro Pro Gln Pro Pro His Pro Pro Leu Ala
675 680 685

```
Tyr Gly Pro Ala Pro Ser Thr Arg Pro Met Gly Pro Gln Ala Ala Pro
    690                 695                 700

Leu Thr Ile Arg Gly Pro Ser Ser Ala Gly Gln Ser Thr Pro Ser Pro
705                 710                 715                 720

His Leu Val Pro Ser Pro Ala Pro Ser Pro Gly Pro Gly Pro Val Pro
                725                 730                 735

Pro Arg Pro Pro Ala Ala Glu Pro Pro Pro Cys Leu Arg Arg Gly Ala
            740                 745                 750

Ala Ala Ala Asp Leu Leu Ser Ser Ser Pro Glu Ser Gln His Gly Gly
        755                 760                 765

Thr Gln Ser Pro Gly Gly Gly Gln Pro Leu Leu Gln Pro Thr Lys Val
    770                 775                 780

Asp Ala Ala Glu Gly Arg Arg Pro Gln Ala Leu Arg Leu Ile Glu Arg
785                 790                 795                 800

Asp Pro Tyr Glu His Pro Glu Arg Leu Arg Gln Leu Gln Gln Glu Leu
                805                 810                 815

Glu Ala Phe Arg Gly Gln Leu Gly Asp Val Gly Ala Leu Asp Thr Val
            820                 825                 830

Trp Arg Glu Leu Gln Asp Ala Gln Glu His Asp Ala Arg Gly Arg Ser
        835                 840                 845

Ile Ala Ile Ala Arg Cys Tyr Ser Leu Lys Asn Arg His Gln Asp Val
    850                 855                 860

Met Pro Tyr Asp Ser Asn Arg Val Val Leu Arg Ser Gly Lys Asp Asp
865                 870                 875                 880

Tyr Ile Asn Ala Ser Cys Val Glu Gly Leu Ser Pro Tyr Cys Pro Pro
                885                 890                 895

Leu Val Ala Thr Gln Ala Pro Leu Pro Gly Thr Ala Ala Asp Phe Trp
            900                 905                 910

Leu Met Val His Glu Gln Lys Val Ser Val Ile Val Met Leu Val Ser
        915                 920                 925

Glu Ala Glu Met Glu Lys Gln Lys Val Ala Arg Tyr Phe Pro Thr Glu
    930                 935                 940

Arg Gly Gln Pro Met Val His Gly Ala Leu Ser Leu Ala Leu Ser Ser
945                 950                 955                 960

Val Arg Ser Thr Glu Thr His Val Glu Arg Val Leu Ser Leu Gln Phe
                965                 970                 975

Arg Asp Gln Ser Leu Lys Arg Ser Leu Val His Leu His Phe Pro Thr
            980                 985                 990

Trp Pro Glu Leu Gly Leu Pro Asp Ser Pro Ser Asn Leu Leu Arg Phe
        995                1000                1005

Ile Gln Glu Val His Ala His Tyr Leu His Gln Arg Pro Leu His Thr
   1010                1015                1020
```

Pro Ile Ile Val His Cys Ser Ser Gly Val Gly Arg Thr Gly Ala Phe
1025 1030 1035 1040

Ala Leu Leu Tyr Ala Ala Val Gln Glu Val Glu Ala Gly Asn Gly Ile
1045 1050 1055

Pro Glu Leu Pro Gln Leu Val Arg Arg Met Arg Gln Gln Arg Lys His
1060 1065 1070

Met Leu Gln Glu Lys Leu His Leu Arg Phe Cys Tyr Glu Ala Val Val
1075 1080 1085

Arg His Val Glu Gln Val Leu Gln Arg His Gly Val Pro Pro Pro Cys
1090 1095 1100

Lys Pro Leu Ala Ser Ala Ser Ile Ser Gln Lys Asn His Leu Pro Gln
1105 1110 1115 1120

Asp Ser Gln Asp Leu Val Leu Gly Gly Asp Val Pro Ile Ser Ser Ile
1125 1130 1135

Gln Ala Thr Ile Ala Lys Leu Ser Ile Arg Pro Pro Gly Gly Leu Glu
1140 1145 1150

Ser Pro Val Ala Ser Leu Pro Gly Pro Ala Glu Pro Pro Gly Leu Pro
1155 1160 1165

Pro Ala Ser Leu Pro Glu Ser Thr Pro Ile Pro Ser Ser Ser Pro Pro
1170 1175 1180

Pro Leu Ser Ser Pro Leu Pro Glu Ala Pro Gln Pro Lys Glu Glu Pro
1185 1190 1195 1200

Pro Val Pro Glu Ala Pro Ser Ser Gly Pro Pro Ser Ser Ser Leu Glu
1205 1210 1215

Leu Leu Ala Ser Leu Thr Pro Glu Ala Phe Ser Leu Asp Ser Ser Leu
1220 1225 1230

Arg Gly Lys Gln Arg Met Ser Lys His Asn Phe Leu Gln Ala His Asn
1235 1240 1245

Gly Gln Gly Leu Arg Ala Thr Arg Pro Ser Asp Asp Pro Leu Ser Leu
1250 1255 1260

Leu Asp Pro Leu Trp Thr Leu Asn Lys Thr
1265 1270

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 493 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Thr Arg Ala Leu Cys Ser Ala Leu Arg Gln Ala Leu Leu Leu Leu
1 5 10 15

```
Ala Ala Ala Ala Glu Leu Ser Pro Gly Leu Lys Cys Val Cys Leu Leu
            20                  25                  30

Cys Asp Ser Ser Asn Phe Thr Cys Gln Thr Glu Gly Ala Cys Trp Ala
        35                  40                  45

Ser Val Met Leu Thr Asn Gly Lys Glu Gln Val Ile Lys Ser Cys Val
    50                  55                  60

Ser Leu Pro Glu Leu Asn Ala Gln Val Phe Cys His Ser Ser Asn Asn
65                  70                  75                  80

Val Thr Lys Thr Glu Cys Cys Phe Thr Asp Phe Cys Asn Asn Ile Thr
                85                  90                  95

Leu His Leu Pro Thr Ala Ser Pro Asn Ala Pro Lys Leu Gly Pro Met
            100                 105                 110

Glu Leu Ala Ile Ile Ile Thr Val Pro Val Cys Leu Leu Ser Ile Ala
        115                 120                 125

Ala Met Leu Thr Val Trp Ala Cys Gln Gly Arg Gln Cys Ser Tyr Arg
    130                 135                 140

Lys Lys Lys Arg Pro Asn Val Glu Glu Pro Leu Ser Glu Cys Asn Leu
145                 150                 155                 160

Val Asn Ala Gly Lys Thr Leu Lys Asp Leu Ile Tyr Asp Val Thr Ala
                165                 170                 175

Ser Gly Ser Gly Ser Gly Leu Pro Leu Leu Val Gln Arg Thr Ile Ala
            180                 185                 190

Arg Thr Ile Val Leu Gln Glu Ile Val Gly Lys Gly Arg Phe Gly Glu
        195                 200                 205

Val Trp His Gly Arg Trp Cys Gly Glu Asp Val Ala Val Lys Ile Phe
    210                 215                 220

Ser Ser Arg Asp Glu Arg Ser Trp Phe Arg Glu Ala Glu Ile Tyr Gln
225                 230                 235                 240

Thr Val Met Leu Arg His Glu Asn Ile Leu Gly Phe Ile Ala Ala Asp
                245                 250                 255

Asn Lys Asp Asn Gly Thr Trp Thr Gln Leu Trp Leu Val Ser Glu Tyr
            260                 265                 270

His Glu Gln Gly Ser Leu Tyr Asp Tyr Leu Asn Arg Asn Ile Val Thr
        275                 280                 285

Val Ala Gly Met Ile Lys Leu Ala Leu Ser Ile Ala Ser Gly Leu Ala
    290                 295                 300

His Leu His Met Glu Ile Val Gly Thr Gln Gly Lys Pro Ala Ile Ala
305                 310                 315                 320

His Arg Asp Ile Lys Ser Lys Asn Ile Leu Val Lys Lys Cys Glu Thr
                325                 330                 335

Cys Ala Ile Ala Asp Leu Gly Leu Ala Val Lys His Asp Ser Ile Leu
            340                 345                 350
```

```
Asn Thr Ile Asp Ile Pro Gln Asn Pro Lys Val Gly Thr Lys Arg Tyr
        355                 360                 365

Met Ala Pro Glu Met Leu Asp Asp Thr Met Asn Val Asn Ile Phe Glu
    370                 375                 380

Ser Phe Lys Arg Ala Asp Ile Tyr Ser Val Gly Leu Val Tyr Trp Glu
385                 390                 395                 400

Ile Ala Arg Arg Cys Ser Val Gly Gly Ile Val Glu Glu Tyr Gln Leu
                405                 410                 415

Pro Tyr Tyr Asp Met Val Pro Ser Asp Pro Ser Ile Glu Glu Met Arg
            420                 425                 430

Lys Val Val Cys Asp Gln Lys Phe Arg Pro Ser Ile Pro Asn Gln Trp
        435                 440                 445

Gln Ser Cys Glu Ala Leu Arg Val Met Gly Arg Ile Met Arg Glu Cys
    450                 455                 460

Trp Tyr Ala Asn Gly Ala Ala Arg Leu Thr Ala Leu Arg Ile Lys Lys
465                 470                 475                 480

Thr Ile Ser Gln Leu Cys Val Lys Glu Asp Cys Lys Ala
                485                 490
```

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	23 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ix) FEATURE:

(D) OTHER INFORMATION: The letter "Y" stands for C or T. The letter "V" stands for A, C or G. The letter "R" stands for A or G. The letter "N" stands for A, C, G or T.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GAYTTYTGGV RNATGRTNTG GGA 23

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	23 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ix) FEATURE:

(D) OTHER INFORMATION: The letter "S" stands for C or G.
The letter "Y" stands for C or T.
The letter "N" stands for A, C, G or T.
The letter "W" stands for A or T.
The letter "R" stands for A or G.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

CGGCCSAYNC CNGCNSWRCA RTG 23

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in positions 4 and 6 stand for an unspecified amino acid.
"Xaa" in position 8 stands for either Glu or Asp.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Asp Phe Trp Xaa Met Xaa Trp Xaa
1 5

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in positions 3 and 6 stand for an unspecified amino acid.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

His Cys Xaa Ala Gly Xaa Gly
1 5

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	34 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CACCGTTCGA GTATTTCAGA TTGTGAAGAA GTCC 34

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	34 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGACTTCTTC ACAATCTGAA ATACTCGAAC GGTG 34

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	33 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CCGTTATGTG AGGAAGAGCC ACATTACAGG ACC 33

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	33 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GGTCCTGTAA TGTGGCTCTT CCTCACATAA CGG 33

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GGCATGCATG GAGTATGAAA TGG 23

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CGTACATCCC AGATGAGCTC AAGAATAGGG 30

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Ser Trp Pro Pro Ser Gly Thr Ser Ser Lys Met Ser Leu Asp Asp Leu
1 5 10 15

Pro Glu Lys Gln Asp Gly Thr Val Phe Pro Ser Ser Leu Leu Pro
20 25 30

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Tyr Ser Leu Pro Tyr Asp Ser Lys His Gln Ile Arg Asn Ala Ser Asn
1 5 10 15

Val Lys His His Asp Ser Ser Ala Leu Gly Val Tyr Ser Tyr
20 25 30

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	30 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

His Thr Leu Gln Ala Asp Ser Tyr Ser Pro Asn Leu Pro Lys Ser Thr
1 5 10 15

Thr Lys Ala Ala Lys Met Met Asn Gln Gln Arg Thr Lys Cys
20 25 30

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	21 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ix) FEATURE:

(D) OTHER INFORMATION: The letter "N" stands for A, C, G or T.
The letter "R" stands for A or G.
The letter "Y" stands for C or T.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GGNCARTTYG GNGANGTNTG G 21

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	24 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ix) FEATURE:

(D) OTHER INFORMATION: The letter "N" stands for A, C, G or T.
The letter "Y" stands for C or T.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CAGNGCNGCY TCNGGNGCNG TCCA 24

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	7 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in position 5 stands for either Glu or Asp.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

Gly Gln Phe Gly Xaa Val Trp
1 5

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	8 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Trp Thr Ala Pro Glu Ala Leu Leu
1 5

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	19 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

AGTGAGCTTC ATGTTGGCT 19

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	18 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

GGTAGAGGCT GCCATCAG 18

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	19 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ix) FEATURE:

(D) OTHER INFORMATION: The letter "N" stands for deoxythymidylate.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GACGATCGGA ATTCGCGAN 19

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	18 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GACGATCGGA ATTCGCGA 18

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	17 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

CCCAGCCACA GGCCTTC 17

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCACACCTCC CCAAAGTA 18

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

TGGGAGCGGC CACACTCCGA ATTCGCCCTT 30

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GCCTGCGTGC GAAGATG 17

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

CTTCGAGGGC ACAGAGCC 18

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

ATGGAGCCGT TCCTCAGGAG G 21

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	21 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

TCACCCAGCT TCCTCCCAAG G 21

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	21 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

AGGCCAACTG GAAGCTGATC C 21

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	20 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GCTGGAGCCC AGAGCGTTGG 20

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	8 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in position 6 stands for an unspecified amino acid.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

His Arg Asp Leu Arg Xaa Ala Asn
1 5

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in positions 6 stands for an unspecified amino acid.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

His Arg Asp Leu Ala Xaa Arg Asn
1 5

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

TCGCCAAGGA GATCCAGACA C 21

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

GAAGTCAGCC ACCTTGCAGG C 21

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	13 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

GGATCCCCGG ACC 13

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	10 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

Met Arg Gly Ser His His His His His His
1 5 10

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	30 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

ATGAGAGGAT CGCATCACCA TCACCATCAC 30

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	8 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in positions 4 and 6 stand for an unspecified amino acid. "Xaa" in position 8 stands for either Glu or Asp.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

Asp Phe Trp Xaa Met Xaa Trp Xaa
1 5

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

His Cys Ser Ala Gly
1 5

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

Met Ser Ser Pro Arg Lys Val Arg Gly Lys Thr Gly Arg Asp Asn Asp
1 5 10 15

Glu Glu Glu Gly Asn Ser Gly Asn Leu Asn Leu Arg Asn
20 25

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

Ser Pro Val Leu Ser Gly Ser Ser Arg Leu Ser Lys Asp Thr Glu Thr
1 5 10 15

Ser Val Ser Glu Lys Glu Leu Thr Gln Leu Ala Gln Ile
20 25

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

Trp Asp Val Ser Asp Arg Ser Leu Arg Asn Arg Trp Asn Ser Met Asp
1 5 10 15

Ser Glu Thr Ala Gly Pro Ser Lys Thr Val Ser Pro Val
20 25

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: The letter "Y" stands for C or T.
The letter "H" stands for A, C or T.
The letter "M" stands for A or C.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

ATCCCCGGCT CTGAYTAYAT HMAYGC 26

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in position 8 stands for either Asn or His.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

Ile Pro Gly Ser Asp Tyr Ile Xaa Ala
1 5

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

Met Glu Glu Leu Gln Asp Tyr Glu Asp Met Met Glu Glu Asn
1 5 10

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

Tyr Gln Gln Gly Gln Asn Gln Leu Tyr Asn Glu Leu Asn Leu Gly Arg
1 5 10 15

Arg Glu Glu Tyr Asp Val Leu Asp Lys Arg Arg Gly Arg Asp
20 25 30

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	8 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in positions 6 and 7 stand for an unspecified amino acid.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

His Arg Asp Leu Lys Xaa Xaa Asn
1 5

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	23 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ix) FEATURE:

(D) OTHER INFORMATION: The letter "R" stands for A or G. The letter "N" stands for Inosine.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:
GARRARGTNG CNGTNAARRT NTT 23

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	29 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ix) FEATURE:

(D) OTHER INFORMATION: The letter "R" stands for A or G.
The letter "N" stands for Inosine.
The letter "K" stands for G or T.
The letter "M" stands for A or C.
The letter "Y" stands for C or T.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

TTRATRTCNC KRTGNGMNAT NGMNGGYTT 29

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	8 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in position 2 stands for Lys or Glu. "Xaa" in position 7 stands for Val or Ile.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

Glu Xaa Val Ala Val Lys Xaa Phe
1 5

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	10 amino acids
(B)	TYPE:	amino acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in position 3 stands for Ala or Ser. "Xaa" in position 5 stands for Ala or Ser.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

Lys Pro Xaa Ile Xaa His Arg Asp Ile Lys
1 5 10

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	24 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

AACTTTGGCT GGTATCTGAA TATC 24

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	24 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

CCTTGTGTAC CAACAATCTC CATA 24

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	22 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

CTCCAGAGAT GAGAGATCTT GG 22

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	22 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

TTCCAGCCAC GGTCACTATG TT 22

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	48 base pairs
(B)	TYPE:	nucleic acid
(C)	STRANDEDNESS:	single
(D)	TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

CTTCGAAAGC TTGAAATCGG TACCATCGAT TCTAGAGTTA ACTTCGAA 48

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH:	47 base pairs
(B)	TYPE:	nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

CTCTAGAACG CGTTAAGGCG CGCCAATATC GATGAATTCT TCGAAGC 47

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

His Cys Ser Ser Gly
1 5

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:
Tyr Arg Lys Lys Lys Arg Pro Asn Val Glu Glu Pro Leu

Claims

1. An isolated, enriched or purified nucleic acid molecule encoding a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide.

2. The nucleic acid molecule of claim 1

(a) having the sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8;

(b) that hybridizes under highly stringent conditions to the sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8;

(c) that encodes a polypeptide having the amino acid sequence set forth in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, or SEQ ID NO:16; or

(d) that encodes a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 domain polypeptide.

3. The nucleic acid molecule of claim 1 where the nucleic acid molecule is isolated, enriched, or purified from a human.

4. The nucleic acid molecule of claim 2 wherein said molecule comprises a

(a) nucleotide sequence that encodes a polypeptide having the amino acid sequence of SEQ ID NO:9 and is lacking at least one of the following segments of amino acid residues: 1 - 48, 49 - 294, 295 - 807;

(b) nucleotide sequence that encodes a polypeptide having the amino acid sequence of SEQ ID NO:10 and is lacking

at least one of the following segments of amino acid residues: 1 - 55, 56 - 109, 120 - 212, 230 - 480, 481 - 488;

(c) nucleotide sequence that encodes a polypeptide having the amino acid sequence of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13 or SEQ. ID. NO:14, and is lacking at least one of the following segments of amino acid residues;

(d) nucleotide sequence that encodes a polypeptide having the amino acid sequence of SEQ. ID. NO: 15 and is lacking at least one of the following segments of amino acid residues: 1 - 857, 353 - 777, 858 - 1096, 1097 - 1274, 1101 - 1214;

(e) encodes a polypeptide having the amino acid sequence of SEQ ID NO:16 and lacking at least one of the following segments of amino acid residues: 1-25, 26-113, 114-493, 193-483; or

(f) hybridizes under stringent conditions to the nucleotide sequence of (a)-(f).

5. A nucleic acid probe for the detection of nucleic acid encoding a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide in a sample.

6. The probe of claim 5, wherein said polypeptide comprises at least 6 contiguous amino acids of the amino acid sequence shown in SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, or SEQ ID NO:16.

7. A isolated, enriched or purified nucleic acid sequence encoding a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 domain polypeptide.

8. A recombinant nucleic acid molecule encoding a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide or a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 domain polypeptide, and a vector or promoter effective to initiate transcription in a host cell.

9. A recombinant nucleic acid molecule encoding

(a) a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide; or

(b) a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 domain polypeptide fused to a heterologous polypeptide.

10. A recombinant cell comprising a nucleic acid molecule encoding

(a) a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide;

(b) a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 domain polypeptide; or

(c) a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide or PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 domain polypeptide fused to a heterologous polypeptide.

11. An isolated, enriched or purified PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide.

12. An isolated, enriched or purified PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 domain polypeptide.

13. The PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 polypeptide of claim 11 wherein said polypeptide comprises

(a) an amino acid sequence of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, or SEQ ID NO:16;

(b) an amino acid sequence encoded by a nucleic acid molecule that hybridizes under highly stringent conditions to the nucleic acid molecule of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8; or

(c) a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 domain polypeptide.

14. An antibody having specific binding affinity to a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide or a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 domain polypeptide.

15. A hybridoma which produces an antibody having specific binding affinity to a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide.

16. A method for identifying a substance capable of modulating PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity comprising the steps of:

(a) contacting a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide with a test substance, and

(b) determining whether said substance alters the activity of said polypeptide.

17. A method for identifying a substance capable of modulating PTP04, SAD, PTP05, PTP10, ALP, or ALK-7 activity in a cell comprising the steps of:

(a) expressing a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide in a cell,

(b) adding a test substance to said cells, and

(c) monitoring a change in cell phenotype, cell proliferation, cell differentiation, PTP04, SAD, PTP05, PTP10,

ALP, or ALK-7 catalytic activity, or the interaction between a PTP04, a SAD, a PTP05, a PTP10, an ALP, or an ALK-7 polypeptide and a natural binding partner.

18. A method of preventing or treating an abnormal condition by administering to a patient in need of such treatment a compound that modulates the function of a PTP04, a SAD, a PTP05, a PTP10, or an ALP polypeptide in vitro.

19. The method of claim 18, wherein said abnormal condition involves abnormality in PTP04, SAD, PTP05, PTP10, or ALP signal transduction pathway.

20. The method of claim 19, wherein said abnormal condition is cancer.

21. A method of promoting neuronal survival by administering to a patient in need of such treatment a substance which modulates an activity of ALK-7 in vitro.

22. A method for identifying modulators of protein activity comprising the steps of:

a) contacting a protein with a natural binding partner, thereby forming a captured protein;

b) contacting said captured protein with a test compound;

c) measuring said protein activity; and

d) comparing said protein activity with the activity of a control protein to determine the extent of modulation, wherein said control protein has the same amino acid sequence of the protein of step a) without said natural binding partner.

23. The method of claim 22, wherein said method utilizes non-radioactive reagents.

24. The method of claim 23, wherein said protein is not a fusion protein.

25. The method of claim 24, wherein said protein is not a GST-fusion protein.

26. The method of claim 25, wherein said protein is an enzyme, a receptor enzyme, or a non-receptor enzyme.

27. The method of claim 26, wherein said protein is a protein kinase.

28. The method of claim 27, wherein said protein kinase is a protein tyrosine kinase.

29. The method of claim 28, wherein said protein tyrosine kinase is Zap70 or Syk.

30. The method of claim 26, wherein said protein is a protein phosphatase.

31. The method of claim 30, wherein said protein phosphatase is a protein tyrosine phosphatase.

32. The method of claim 31, wherein said protein tyrosine phosphatase is PTP04, SAD, PTP05, PTP10, ALP, or ALK-7.

33. The method of claim 23, wherein said natural binding partner is capable of binding to a solid support.

34. The method of claim 33, wherein said natural binding partner is a peptide.

35. The method of claim 34, wherein said peptide comprises a phosphopeptide.

36. The method of claim 35, wherein said phosphopeptide comprises an ITAM motif.

37. The method of claim 33, wherein said natural binding partner comprises a lipid.

38. The method of claim 33, wherein said solid support comprises well plate, glass beads, or resin.

39. The method of claim 23, wherein said activity is autocatalytic activity, catalytic turnover of substrate, or binding of a second natural binding partner.

40. The method of claim 23, further comprising the step of contacting said capture protein with one or more components of the group consisting of a substrate, a second natural binding partner, and an antibody.

41. The method of claim 23, further comprising the step of lysing cells containing said protein prior to step (a).

42. A kit for the identification of modulators of non-receptor enzyme activity comprising:

a) a natural binding partner;

b) a solid support; and

c) a binding agent.

43. The kit of claim 42, wherein said binding agent is selected from the group consisting of a substrate, a second natural binding partner, and an antibody.

44. The kit of claim 43, wherein said natural binding partner is a peptide.

45. The kit of claim 44, wherein said peptide is a phosphopeptide.

46. The kit of claim 45, wherein said phosphopeptide comprises an ITAM motif.

47. The kit of claim 42, wherein said natural binding partner comprises a lipid.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/54, 15/55, 9/12, 9/16, C07K 14/705, C12N 15/11, C07K 16/40, 16/28, C12N 5/12, 15/62, C12Q 1/42, 1/48

A3

(11) International Publication Number: WO 98/493

(43) International Publication Date: 5 November 1998 (05.11.

(21) International Application Number: PCT/US98/08439

(22) International Filing Date: 27 April 1998 (27.04.98)

(30) Priority Data:

60/044,428	28 April 1997 (28.04.97)	US
60/047,222	20 May 1997 (20.05.97)	US
60/049,756	11 June 1997 (11.06.97)	US
60/049,477	11 June 1997 (11.06.97)	US
60/049,914	18 June 1997 (18.06.97)	US
60/063,595	23 October 1997 (23.10.97)	US

(71) Applicant *(for all designated States except US)*: SUGEN, INC. [US/US]; 351 Galveston Drive, Redwood City, CA 94063 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: PLOWMAN, Greg, D. [US/US]; 4 Honeysuckle Lane, San Carlos, CA 94070 (US). CLARY, Douglas [US/US]; 164 Midcrest Way, San Francisco, CA 94131 (US). JALLAL, Bahija [MA/US]; 101 O'Keefe Street, Melo Park, CA 94025 (US). PELES, Elior [IL/IL]; Hanasi Harishon 51, 76303 Rhovot (IL). ONRUST, Susan [US/NZ]; 6 Summit Drive, Mt. Albert, Auckland 3 (NZ). MARKBY, Dave [US/US]; Apartment A, 477 Burnett Avenue, San Francisco, CA 94131 (US). COURTNEIDG Sara, A. [GB/US]; 1408 Alvarado Avenue, Burlingame, 94010 (US). APP, Harald [DE/US]; 630 27th Street, Francisco, CA 94131 (US). HUI, Terance, H. [CN/US]; Skyline Drive, Daly City, CA 94015 (US).

(74) Agents: WARBURG, Richard, J. et al.; Lyon & Lyon LL Suite 4700, 633 West Fifth Street, Los Angeles, 90071-2066 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARI patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Euras patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), Europ patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report:
25 February 1999 (25.02.

(54) Title: DIAGNOSIS AND TREATMENT OF PHOSPHATASE OR KINASE-RELATED DISORDERS

(57)Abstract

The present invention relates to phosphatases and kinases, nucleic acids encoding such polypeptides, cells, tissues and anin containing such nucleic acids, antibodies to such polypeptides, assays utilizing such polypeptides, and methods relating to all of foregoing. Methods for treatment, diagnosis, and screening are provided for phosphatase- or kinase-related diseases or conditi characterized by an abnormal interaction between a phosphatase or a kinase and its binding partner.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the I

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

Interna[illegible]nal Application No
PCT/US 98/08439

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/54 C12N15/55 C12N9/12 C12N9/16 C07K14/705
C12N15/11 C07K16/40 C07K16/28 C12N5/12 C12N15/62
C12Q1/42 C12Q1/48

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	WO 96 34985 A (SUGEN INC) 7 November 1996 see page 5, line 3 - page 6, line 9 see page 15, line 17 - line 24 ---	5,6 1-4,7-17
X	EMBL database entry HS1185621; accessionnumber AA281242; 4. April 1997; Robert Strausberg: 'National Cancer Institute, Cancer Genome Anatomy Project.' XP002076843 see abstract ---	2,4-6
A	WO 95 06735 A (LUDWIG INST CANCER RES ;GONEZ LEONEL JORGE (SE); SARAS JAN (SE); C) 9 March 1995 see page 43, line 5 - page 44, line 25; examples 1-8 --- -/--	1-20, 22-26, 30-41

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 September 1998	21. 12. 98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Mandl, B

8

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte . tional Application No
PCT/US 98/08439

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	AOKI N. ET AL.: "THE NOVEL PROTEIN-TYROSINE PHOSPHATASE PTP20 IS A POSITIVE REGULATOR OF PC12 CELL NEURONAL DIFFERENTIATION" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 271, no. 46, 15 November 1996, pages 29422-29426, XP002053901 see 'Experimental Procedures' ---	1-17
A	SARAS J. ET AL.: "CLONING AND CHARACTERIZATION OF PTPL1, A PROTEIN TYROSINE PHOSPHATASE WITH SIMILARITIES TO CYTOSKELETAL-ASSOCIATED PROTEINS." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 269, no. 39, 1994, pages 24082-24089, XP002076840 see the whole document, especially 'Materials and Methods' ---	1-17
A	MATTHEWS R. J. ET AL.: "Characterization of hematopoietic intracellular protein tyrosine phosphatases: Description of a phosphatase containing an SH2 domain and another enriched in proline-, glutamic acid-, serine-, and threonine-rich sequences." MOLECULAR AND CELLULAR BIOLOGY, vol. 12, no. 5, 1992, pages 2396-2405, XP002076841 see the whole document ---	1-17
A	MAEKAWA K. ET AL.: "MOLECULAR CLONING OF A NOVEL PROTEIN-TYROSINE PHOSPHATASE CONTAINING A MEMBRANE-BINDING DOMAIN AND glgf REPEATS." FEBS LETTERS, vol. 337, 1994, pages 200-206, XP002076842 see 'Materials and Methods' ---	1-17
A	WO 93 10242 A (COLD SPRING HARBOR LAB) 27 May 1993 see page 6, line 21 - page 7, line 23 -----	22-26, 30-41

8

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

compound that modulates the function of a PTP05 tyrosine phosphatase; a method of identifying modulators of protein activity as far as they refer to a PTP05 tyrosine phosphatase.

4. Claims: 1-20, 22-26 and 30-41 (all partially)

A PTP10 tyrosine phosphatase and the nucleic acid encoding it; a nucleic acid probe for the detection of said nucleic acid; a nucleic acid molecule encoding a PTP10 tyrosine phosphatase and a vector or a promoter; a nucleic acid molecule encoding a PTP10 tyrosine phosphatase fused to a heterologous polypeptide; a recombinant cell comprising a nucleic acid encoding a PTP10 tyrosine phosphatase; an antibody specific for said PTP10 tyrosine phosphatase and a hybridoma cell that produces said antibody; a method for identifying a substance capable of modulating the activity of the PTP10 tyrosine phosphatase; a method for preventing or treating an abnormal condition by administering a compound that modulates the function of a PTP10 tyrosine phosphatase; a method of identifying modulators of protein activity as far as they refer to a PTP10 tyrosine phosphatase.

5. Claims: 1-20,22-26 and 30-41 (all partially)

An ALP tyrosine phosphatase and the nucleic acid encoding it; a nucleic acid probe for the detection of said nucleic acid; a nucleic acid molecule encoding an ALP tyrosine phosphatase and a vector or a promoter; a nucleic acid molecule encoding an ALP tyrosine phosphatase fused to a heterologous polypeptide; a recombinant cell comprising a nucleic acid encoding an ALP tyrosine phosphatase; an antibody specific for said ALP tyrosine phosphatase and a hybridoma cell that produces said antibody; a method for identifying a substance capable of modulating the activity of the ALP tyrosine phosphatase; a method for preventing or treating an abnormal condition by administering a compound that modulates the function of an ALP tyrosine phosphatase; a method of identifying modulators of protein activity as far as they refer to a PTP10 tyrosine phosphatase.

6. Claims: 21 (complete) and 1-20, 22-27, 33-41 (partially)

An ALK-7 type I receptor ser/thr kinase and the nucleic acid encoding it; a nucleic acid probe for the detection of said nucleic acid; a nucleic acid molecule encoding an ALK-7 type I receptor ser/thr kinase and a vector or a promoter; a nucleic acid molecule encoding an ALK-7 type I receptor ser/thr kinase fused to a heterologous polypeptide; a recombinant cell comprising a nucleic acid encoding an ALK-7 type I receptor ser/thr kinase; an antibody specific for

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

said ALK-7 type I receptor ser/thr kinase and a hybridoma cell that produces said antibody; a method for identifying a substance capable of modulating the activity of the ALK-7 type I receptor ser/thr kinase; a method of promoting neuronal survival by administering a compound that modulates the activity of an ALK-7 type I receptor ser/thr kinase; a method of identifying modulators of protein activity as far as they refer to an ALK-7 type I receptor ser/thr kinase.

7. Claims: 29,42-47 (complete) and 22-28,30-41 (partially)

A method for identifying modulators of protein activation as far as they do not refer to the following phosphatases or kinases: PTP04, PTP05, PTP10, SAD, ALP and ALK-7; and a kit for the identification of modulators of non-receptor enzyme activity.

INTERNATIONAL SEARCH REPORT

Information on patent family members

In.. .ational Application No
PCT/US 98/08439

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9634985 A	07-11-1996	US 5830648 A	03-11-1998
		AU 5636996 A	21-11-1996
		CA 2220300 A	07-11-1996
		EP 0824597 A	25-02-1998
WO 9506735 A	09-03-1995	AU 683299 B	06-11-1997
		AU 7644394 A	22-03-1995
		CA 2170515 A	09-03-1995
		EP 0789771 A	20-08-1997
		JP 9510861 T	04-11-1997
		NZ 273219 A	26-06-1998
		US 5821075 A	13-10-1998
WO 9310242 A	27-05-1993	US 5294538 A	15-03-1994
		CA 2122338 A	27-05-1993
		EP 0614489 A	14-09-1994
		JP 7503131 T	06-04-1995
		US 5441880 A	15-08-1995
		US 5695950 A	09-12-1997
		US 5770423 A	23-06-1998
		US 5756335 A	26-05-1994